DOCUMENT RESUME

ED 073 377 CG 007 785

AUTHOR Hodge, Francis P., Ed.
TITLE A Workshop with Donald P. Hoyt: Uses of Appraisal and Measurement for the Improvement of Vocational Counseling.
INSTITUTION State Univ. of New York, Albany. Two Year Coll. Student Development Center.
PUB DATE 72
NOTE 45p.; Paper presented at Institute on Man and Science (Rensselaerville, New York, May 10-12, 1970)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS College Students; Conference Reports; Evaluation Methods; *Evaluation Techniques; Higher Education; *Learning Experience; Learning Processes; *Measurement Techniques; Student Characteristics; Student Development; *Student Evaluation; Symposia; *Vocational Counseling; Workshops

ABSTRACT

This workshop report focuses on the necessity of broadening the base of evaluation in higher education and beginning to look at types of student development other than purely academic: their characteristics, strengths, weaknesses, goals, values, motivations, and most effective learning styles. The need for non-standardized measures occurs in two contexts. First, in helping students plan their future there is frequently a need for appraisal of special talents or inclinations beyond those concerned with academic background or potential. Second, is trying to determine the effectiveness of a given program--such as counseling method, teaching approach, or orientation program--standardized measures are seldom appropriate indicators of success. The workshop provided some principles underlying non-standardized appraisal techniques. The two principles essential for this individual, primarily non-academic appraisal are 1.) student self-appraisal, asking "him" what you want to know about "him," and 2.) prediction of future behavior by past behavior which resembles the future behavior. Both principles, the workshop leader felt, suggest the need for awareness of, or determination of, the environment where for a given individual learning experiences have most likely occurred. Practical applications of these principles are included as examples from the workshop. References are included. (Author/SES)

ED 073377

WORKSHOP REPORTS

A WORKSHOP WITH DONALD P. HOYT: USES OF APPRAISAL AND MEASUREMENT FOR THE IMPROVEMENT OF VOCATIONAL COUNSELING

Edited by: FRANCIS P. HODGE

FILMED FROM BEST AVAILABLE COPY

SPONSORED BY:

TWO-YEAR COLLEGE STUDENT DEVELOPMENT CENTER
STATE UNIVERSITY OF NEW YORK AT ALBANY
1400 WASHINGTON AVENUE - EDUCATION 314
ALBANY, NEW YORK 12222

WILLIAM A. ROBBINS, DIRECTOR

1972

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

CG 007 185

ED 073377

A WORKSHOP WITH DONALD P. HOYT:

USES OF APPRAISAL AND MEASUREMENT FOR THE IMPROVEMENT OF VOCATIONAL COUNSELING

EDITED BY FRANCIS P. HODGE

May 10-12, 1970

Institute on Man and Science, Rensselaerville, New York

Two-Year College Student Development Center
State University of New York at Albany
1972

PREFACE

Should we continue to follow traditional testing practices in our community colleges? If we do not, then how can we reform our educational practices to provide us with a more accurate reading of students--their characteristics, strengths, weaknesses, goals, values, motivations, and most effective learning styles? "It is essential that we now begin to broaden our base of evaluation in higher education and start looking at types of student development other than purely academic." In the spring of 1970 this statement by Donald P. Hoyt, Director of the Office of Educational Research, Kansas State University, plus the issues inherent in the above questions, provided the focus for a two-day workshop sponsored by the Two-Year College Student Development Center. During this workshop, Dr. Hoyt, assisted by William Feister, Department of Guidance and Personnel Services, State University of New York at Albany, led the participants through a study of non-traditional approaches to appraisal and measurement. Before coming to SUNY-Albany, Dr. Feister was with Dr. Hoyt at the American College Testing Program.

Our feeling was that formal courses in tests and measurements or in test interpretation were available in most graduate programs. But practical assistance in appraising characteristics for which standardized measures are inadequate or do not exist is seldom available. The workshop was designed to introduce participants to methods which would fill this gap.

The need for non-standardized measures occurs in two contexts. First, in helping students plan their future there is frequently a need for appraisal of special talents or inclinations beyond those concerned with academic background or potential. Second, in trying to determine the effectiveness of a given program (counseling method, teaching approach, orientation program, etc.), standardized measures are seldom appropriate indicators of success. The workshop provided some principles underlying non-standardized appraisal techniques. Participants had an opportunity to review practical applications of these principles, and to make some of their own applications.

Spring 1970 was a turbulent period. I well remember wondering whether or not we should postpone the workshop since some participants were forced to cancel their reservations and deal with campus problems. In the two years that have passed, much has taken place. However, the topics and material still seem pertinent and timely, appropriate reading for personal or staff in-service education. We have therefore reviewed the tape recording, that fortunately was made at the workshop, and fashioned what we hope will be a usable report to the educational field. This report should help workshop participants themselves to refresh their memory of the discussion, but also should stimulate many others in college counseling, teaching and administrative work to carry out pertinent studies.

The reader should remember that this material has been taken directly from tapes. In preparing this material we have deliberately preserved the informal conversational form in which Dr. Hoyt conducted his presentations. Sometimes we have been sorely tempted to turn the material into something like articles or papers. However, we have resisted the temptation and have contented ourselves with editing out duplication, and extraneous remarks. For best usage by the reader, discussions that continued across sessions at the workshop are joined together in a continuous narrative.

As I looked over this material after a gap of time, I was impressed with the balanced, careful briefing that Dr. Hoyt provided the workshop participants. Dr. Hoyt's educational vision, humanness, and perspective on the foibles of our traditional educational process emerge very clearly and speak effectively to those of us in the community college movement. The reader can thus benefit, even though unfortunately he or she cannot sense the give and take of a retreat-type seminar that developed the topics in much more depth and accuracy.

Our workshop setting was an ideal one--the Institute on Man and Science, at Rensselaerville, New York, just a short distance southwest of Albany, New York. In every way the physical facilities of this estate-like conference spot, together with the warmth and understanding of the Institute staff, contributed to our workshop goals.

The workshop was an important link in a series supported by the Vocational Education Act. These workshops were aimed at improving the competencies of counselors trying to help community college students with the array of vocational, educational and personal problems facing them. Dorothy Knoell, Terry O'Banion, Allen Ivey, and many other state and national leaders contributed to make that project a very exciting one. We owe much gratitude to Lawrence Gray and Paul Chakonas, in the Bureau of Two-Year College Programs, New York State Education Department, and to S.V. Martorana in the State University of New York, for their vision and confidence in giving support to this program.

Francis P. Hodge, Assistant Director of the Cente has pulled together ten or more hours of taped discussion without disturbing unnecessarily the conversational character of the material, into a report of relative order and smoothness. Both Donald Hoyt and I (together with many readers, I am sure) are grateful to him for his contribution.

William A. Robbins
Director

Charge to workshop participants at the opening session by William A. Robbins, Director, Two-Year College Student Development Center.

Our concern is students--not techniques, but students--and what we can do to provide help to students in our work and in the work of the two-year colleges and urban centers. Our concern is how we can know more about students in order that we can help them. In our effort to know more about students, we need to employ every possible approach we can take, and then we must turn data and knowledge into programs of assistance and real help. We need to be concerned constantly about the validity of this data and about the ends of our programs, the values, the feasibility, and based on what we know, whether there is payoff in the kinds of programs we are executing. So, we have a doublesided need--to learn more so that we can help, and to survey regularly the programs in which we are engaged. As we often have said to ourselves and have had told to us, the two-year colleges and urban centers are at the cutting edge and are faced with an immense task.

This kind of workshop is a way by which we can become better alert to how we can carry out our work well. Over the next two days Dr. Hoyt will help us find special tools to accomplish this task. I would like to present to you Donald P. Hoyt, Director of Educational Research, Kansas State University...

USES OF APPRAISAL AND MEASUREMENT

FOR THE IMPROVEMENT OF VOCATIONAL COUNSELING

Educational measurement and appraisal constitutes one of the three fundamental processes by which the educational enterprise succeeds or fails. This is one of the three things in education that, if we do well, it will succeed; if we do poorly, it will fail. For the record the other two, aside from measurement and appraisal, are first the provision of educational opportunities or the providing of experiences in an environment which turns out to be truly educative, and secondly, supplying educational facilitators. One has to have the people, the machines, and the materials which insure that the educational experience becomes internalized in some way. Consequently, the facilitators and the opportunities are needed, and the third necessity is some method of measurement and appraisal.

Why should this method of measurement and appraisal play such a key role? There are several things to note. In education we are concerned with the development of individual potentials. The chief way by which we can discover who an individual is and which of his potentials can be developed is through the measurement and appraisal process. If we do not measure or appraise, an understanding of who it is we are dealing with is impossible. Consequently, we have little idea what education ought to mean for that individual. What are his individual talents, inclinations, desires; what characteristics does this individual have which constitute roadblocks to the development or realization of his potentials and his goals? The measurement and appraisal process helps find answers to those questions. Unless these questions are answered, one can have great teachers, great programs, and opportunities but not much else, if individuals do not get to the right program, right teacher, or the right facilitator. So what characteristics constitute roadblocks to any individual realizing his goals is a very important concept. Too often in higher education, in community colleges as well as four year colleges, we assume that everybody has the same goals. Further we assume we can say what the goals are. Most individuals accept this without question. We have been quite wrong on that. If we really are concerned with the development of potentials, we must appraise, or we can be upset about the job of developing those in some reasonable way, through our people, through our programs.

Secondly, as justification for the key role assigned to appraisal in education, we have a responsibility to society to foster the learning process through appraisal of an individual's success. We accept this committment. We inform society as to the status, the progress, the success of a given individual who has been under our guidance. Perhaps, more important than that particular function though, we have the responsibility of fostering learning which, as a result of practice, almost any theorist will state requires reinforcement or reward. People learn those things for which they are reinforced, those things for which they are rewarded or those things about which they feel good. The appraisal of the individual, and the feeding back of information to him, constitute a relatively potent source of reward. Reward, therefore, becomes an integral part of the learning process. The way in which we reinforce, and what we reinforce for, determines in large degree what progress is made with an individual. Success is a key point. What is success? What is educational success? And how is a person appraised if he is doing well or doing poorly?

The third justification for this key role given appraisal is the fact that educational programs exist to promote individual development. Without some sort of assessment, measurement of the impact of these programs in terms of their goals, how are these programs to be improved? Unless appraisal of the success of the program is undertaken, is there any way in which we'll know whether or not a particular program ought to be changed or updated? How will we know what innovations to retain and which ones to discard or what traditions to keep and which ones to change? Will we know how to modify the programs of learning or counseling or other services to students unless we take a look at the program's impact on each individual? What do these various programs do to promote an individual's learning or development?

Almost always the focus of educational measurement and appraisal is on the individual and individual characteristics. However, what is the guideline? We take a look at a brand new student, someone whom we have never seen before. What are we trying to do? We appraise this student. Basically, we try to learn something about his individuality. What are his talents, proclivities, needs, style? What is he? Who is he? How does he differ from the next guy? Presumably this pre-measure gives us some indication of what this individual is like. Note this same pre-measure for other people who in some way or another resemble this student can be put together. For example, all students from little rural schools, or all blacks from ghetto areas, or all students from white suburbia could be put together. These collections constitute sub-cultures. We put together what we learned in our appraisal, and then we can sketch a picture of the sub-culture in which the individual belongs. Our understanding of the sociological, social, psychological environment to which he was exposed gives us a better background for understanding the means of individual characteristics, or the ways in which the student differs or conforms to the particular sub-culture to which he belongs. More than one use can be made of the appraisal of an individual taken at the time he enrolls at an institution.

Post-measure, or some measure of an individual taken after he has been around for a time are available also. These post-measures may be taken not necessarily after the student is finished, but rather anytime after entry or after the institution or programs have influenced him. When this is completed on an individual basis, personal development, rather than individuality, has been appraised. The changes in style, need, talents and proclivities are noted. At this time the development question being investigated requires also appraisal. If this is done in a conglomerate similar to the sub-cultures for the pre-measures, this is called program evaluation. For example, if the post-measures who have been exposed to a curriculum of data processing are examined, because this is a conglomerate of their development, we get some idea of whether the data processing curriculum had any degree of success or not. Examine the people exposed to a program in orientation, or a program in counseling, and some of the same inferences might be made. Therefore, when we examine a conglomerate group of people who have been exposed to a similar program or treatment, program evaluation through the appraisal of individual characteristics results.

By and large measurement and appraisal do not serve education very well. Capitalization on this ingredient in education must overcome several obstacles. First, we must acknowledge that we have had available for our use quite limited measures of initial student characteristics. We have catalogs full of books or tests which concern intelligence, scholastic aptitude, or academic achievement. Sometimes structured personality tests or vocational interest are included. There is little of anything which concerns itself with the variety of characteristics which make the world go around. Nothing about how persuasive an

individual is, or how dependable he might be, or really how motivated or driven he might be is included. Items which present techniques or rules or that mathematical formula lend themselves well to are found in abundance. Some good measures result from this type of treatment, but not everything lends itself well to that, and, as a result, a very limited coverage from which to borrow our assessment of individual student characteristics is available.

An even greater limitation has to do with the fact that we have very limited measures of student development. What we think are measures of interest, or what a student likes or becomes are extremely limited for the most part. The only things available are the GRE's, the GED's or similar types of academic measures. These measures say the student has learned "so much English," or "so much math," or other content. Generally though we do not use this information because we have transcripts which tell us **everything** we have to know right and wrong about a student's development. So, this is a second reason why educational measurement and appraisal, while it holds a potentially key role, has not contributed in insuring that the educational enterprise succeeds, anywhere near the extent possible.

Finally, a few attitudinal obstacles which in the long run are probably the most important, must be recognized. One obstacle alluded to earlier relates to the limited availability of measures. If a characteristic is not on our profile summary sheet, then the characteristic does not exist. If when the student profile is drawn, there is no measure for a given characteristic, then that characteristic simply doesn't exist. Therefore, there is no need to worry about it or concern oneself about anything not included on the summary sheet. This attitude has been taught, partially because in the study of tests and measurement concentration is directed toward what tests and measurements are available. Availability assumes, therefore, an unnecessary importance. Also in part this attitude exists because to trust the profile summary and not worry about other responsibilities is easier.

A corollary to this attitude is that anything worth measuring can be found in ETS or the SRA catalog. We assume these people are measurement specialists who know what measurement is and how to go about it. Therefore, if a characteristic is important, they will have a measurement instrument for it. A third attitude related to the previous two is, "I can't do anything about it." This attitude also presents obstacles. A further obstacle to the use of appraisal for the development of individuals is the defeatist attitude resulting from a lack of advanced study or knowledge of measurement science or mathematics.

Various types of activities can be done to develop more optimistic attitudes about roles and processes in the measurement and appraisal program. To discover what the traditional method includes there are many textbook references which are readily available. This traditional method is not bad, but very limited. Our concern should be to discover new ways, non-traditional ways to broaden the pace of individual assessment.

In the material which follows various concepts and points of view about measurement will be presented. These materials will be illustrated from practical experience and applied to help utilize the appraisal and measurement techniques in a bona fide effort to examine how sound educational decisions are made.

Dr. Hoyt: I want to get some principles outlined for you as to how one would go about determining the success, the relative success or failure, of the counseling program, the orientation program, the curricular program, of any kind of programmed experience which you thought was educational, but which may not be. I hope to have time to tell you how I think that ought to be done. I'll give you some illustrations from experiences used either at my office or other places where I've been.

The most important concern will be to have good questions knowing something about how you might go about developing questions. Also I hope you might get a chance to look briefly at the logic of developing your own study on your own campus to answer those questions put to you everyday. This might give you some basis for making other than a seat of the pants decision as to how you go about answering such questions. The assessment of individual characteristics, an activity extremely useful to advising, counseling and looking at the individual will be examined. A chance to apply this assessment and related understandings to the question of whether or not a program is working well, will be discussed later. Also utilizing the appraisal and measurement techniques in a bona fide effort to look at the making of sound educational decisions will be examined.

Question: One of the things that I came for was perhaps some reinforcement of seat of the pants decisions. I'm concerned about assessment of college students because it's hard to know what kind of students would fit best into a college. So, I hope that I'll get out of this conference measures of assessing humans other than the kind of tests we currently have available.

Dr. Hoyt: Yes, I am trying to focus on ways of getting to know more about students than we learn through structured tests currently available. This is surely a major focus for this workshop. The point about the seat of pants is well taken. Don't misunderstand me --I had to fly by the seat of my pants 80 percent of the time anyway. Presently I don't have the time. However, now and then some big decisions come up. Big decisions about whether there are going to be 12,000 students next year. To make a decision like that, I want to have something more than the seat of my pants. That's all I'm saying. If you've got to fly, fly. But if you have to go blind, go blind. You don't always have to go blind. Progress, though, is slow.

Question: Are there ways for measuring drives that you can present to us?

Dr. Hoyt: I'm not going to talk about that explicitly. Although after we have talked about some of the approaches as to how one would go about this I think we can discuss that. I feel there are some very special problems associated with that concept which would require prior consideration. Maybe we'll have a chance to bring some of those out.

Questions: 1) Concern about reinforcement which should be given to the "C" student as well as the "A" student, and concern about attitudinal measurement and standardized statistics have been expressed. How do we apply a better method of using standardized tests for urban center students, for example?

2) How can a student , germane to his self-perception, appraise himself?

3) How can we turn on faculty and students who are up-tight about testing programs? How can we sell the testing program and support it so the faculty and students will buy it?

4) Have measurement and appraisal instruments currently in use really proved themselves? There's basically a feeling that there has been an overstressing of some of the various instruments in use, and that is sometimes for the convenience of such groups at ACT or Educational Testing Services, etc. This is their business, and they're in business to stay in business.

5) How could we best gain the support of the leadership in education? The presidents on the campus, the deans, the division chairmen, etc., how can we gain their support for effective measurement and appraisal programs?

6) How can the data that is collected be best utilized by faculty to help their students succeed?

Dr. Hoyt: I want to answer all the questions but I feel some are on target and some of them aren't. I'll make just a comment or two about some items I've noted here.

I have a few ideas on reinforcement of the "C" student. You mentioned attitudinal measurement and standardized tests particularly in use in the Urban Center for disadvantaged students. I'm for this; everyone's for this; I don't think I'm alone. I do feel that we should try to appraise and measure. If one is to counsel or to teach, he has to know the background. He has to have some sort of relevant experience. Something about how easily one gets away from this information has been mentioned. Part of a middleclass background which doesn't let you enter freely makes you unable to talk. You can't understand, you can't feel, and I feel very much personally this way. I think if not the most significant, this inability to talk is one of the two most significant problems in education today. I feel quite badly about that, and I want you to know I don't have all the answers.

How can a student appraise himself? I'll have some illustrative material on that. It's a hot idea. I think we really missed the boat on this in the past, and I think it's one of the ways we can go forward rapidly.

Resistance to testing and how to enlist more support for testing. Well, take the word "testing" out and talk about appraisal. That's a very important question. We are going to have to deal with this problem. I will have a few things to say about that later. However, not enough so you can go back and feel confident everyone on the campus is going to think they're the most important person on that campus. For some good reasons this has become a serious educational problem, and we are in danger of throwing the baby out with the bath. This is too bad.

Have measuring instruments really proved themselves, or are they pretty nearly something in the self-interest of the test makers. I know that's a loaded question. I know that they aren't going broke. ACT isn't broke by a long shot, and neither is ETS. They're healthy, and so are most of the testing bureaus. Therefore, they seem to be serving their own interests. If you look at the evidence on whether their tests do what they say they'll do, you'll agree that for the most part, the testing bureaus stand up pretty well. My complaint is on what they say they'll do and how important that is. How much of a story are they really telling? How really comprehensive a focus on educational enterprise are they providing for us? There I do not feel they've done a good job at all. I don't want to be simply critical of those firms. I worked for one for a while. When you work for them, you become very aware

you are the vender and they are the customer. We'll tell you what we want to change, not vice versa. It's tough, but I think it will probably always be this way. They'll serve, but we've got to tell them what it is we want, and what we need. They'll do it. They're smart. They'll go out of business if they don't. It's a dual role. As far as what's needed--what's required, that's going to be the function of the people who are in education, to tell those experts.

On another question I am worried about gaining support of the leaders. To a lot of people these days, testing, in particular, is certainly a clay pigeon, and a pretty good clay pigeon. Without even trying too hard, they make a good case against it, but again I say, I think, the answer is really up to bat. Examine what is currently popular, pass-fail, credit-no credit, or ways of trying to solve some of these problems. I think it's not going to solve anything at all. It's not that appraisal of success is wrong, but we have been so limited in what we've been looking at, what we call success. We look at one little dimension. We say that that's all success is. We say we're going to cure the problem by throwing success appraisal out. However, the problem is to look at all the ways in which students might develop. We need to get some feedback and gain some reasons as to why we need appraisal, rather than to say "Forget about it--appraisal is not important." It is important. Appraisal is important both to society and to the individual. It's the one sure way we've got of reinforcing and best becoming an integral part of the learning process. You discard that, and you throw away much of your potency as learning facilitators.

How can collected data be used by faculty to help students succeed? This is a critical--critical question. You probably won't like to hear this. The one reason I never win in an honest election is I tend to be frank. So, I'll say this to you too. One of your problems is you call yourselves student personnel workers or counselors. You have been proud of this separateness. You identify yourselves as student affairs people, and those are faculty affairs people over there in the classroom. I think this has been one thing that's killed us. In the past this separateness may have been necessary to institute some really important reforms in education in the country and to correct some of the short-sighted philosophies. However, at this time it's strangling education, I think, and it will become more and more important for us to blur the lines between student services and faculty services and to start talking about us all as educational agents. Now the question hardly bears that diatribe, but I wanted to emphasize that we must all consider ourselves educators, that's what we're there for--in education. Yet we all, faculty and counselors, have to learn more about how the educational process proceeds. One important part of this process is appraisal, or knowing who the customer is and knowing how he is proceeding. Finding the opportunities whereby the student can stretch and grow and develop. When we put it that way, we quit thinking about finding the laboratories and finding the classrooms, and we start thinking of finding the opportunities where a student can learn something. These opportunities can be in the community, across the state, around the neighborhood, or at the church. Wherever opportunities are where a student can learn. This is important and is when we start getting at some decent characterizations of ourselves as faculty, administration, student personnel. When those lines start to blur, education starts to occur. <u>WE</u> all need help in interpreting, using, and developing the data to tell us who a student is and how he's progressing. I hope that maybe some of the things we've been saying will be helpful to you and through you to your students and colleagues.

APPRAISAL OF INDIVIDUAL DIFFERENCES

To gain some educational perspective on this topic and why it is important in this workshop, a review of past research is in order. Several types of studies will be mentioned, not in any great detail, but to give a few highlights and to suggest primary implications for where we must go in education.

First, let's take a quick look at summary statistics on characteristics of junior college students. Dorothy Knoell*, Lee Medsker, and Patricia Cross have done a number of these studies and they all report about the same sort of thing. Our concern, though, is not the diversity within junior colleges, but the diversity among junior colleges. When we say, "This is what a junior college is like," are we really talking about the typical junior college or is there such a thing as a typical junior college? At ACT we collected materials on public community colleges--community colleges supported with public funds. One thing we did was examine the average academic ability input, by taking the ACT composite score, the same thing as the SAT, and you see that the averages vary quite widely. The lowest of the samples would place the average student in this junior college at the 14 percentile on a national sample.

At the other extreme, the average student would be at the 70 percentile. What's normal at this college would be at the 70 percentile nationally for college-going students; most of whom go to four-year colleges. At this college, it would be the 14 percentile. I must say that compared to four-year colleges, even that's small especially when you look at the diversity among Ph.D.-granting institutions, and the entering freshmen classes in institutions which grant Ph.D. degrees. When we look at that in terms of equated IQ scores, the average IQ for entering freshmen at level four Ph.D.-granting institutions varies from about 133 to 80. The point is that there is much diversity, and not just in academic input of this type. For example, if you examine what percent of the entering class works part-time, you will find a tremendous range. At one institution--these are again from the public junior colleges--32 percent of the class works part-time. At the other extreme 86 percent works part-time. These figures would indicate some differences between these institutions.

If you asked students what the major reasons were why they came to this college, and you give them one of the preferences, "close to home," you discover at one public junior college, 71 percent said that's the major reason. At another, 5 percent said that's the major reason. There has to be some difference between those colleges.

If you ask students "What's the major goal you have in going to college?" and at one public junior college 66 percent said vocational preparation, while another 33 percent said vocational preparation, it has to make some difference which of these colleges you're working at as to what the nature of your job is, what the nature of the educational task is. There is enormous diversity among junior colleges.

Second, let me describe a few studies which have tried to get at what the nature of academic success is. These are important studies because academic success, as I indicated previously have been about our only measure of success for the college student. I want to give you a brief rundown on this because I think it's good background material for the workshop. At the time when I was learning to be a counselor, it was common to give trainees advice on how to handle one or another situation. The most common advice when you were dealing

* See Bibliography for references indicated throughout this report.

with the student having a rough time academically, maybe in mathematics, was to tell him, "Don't worry, maybe your field is history, or political science or economics. Everybody's got his strengths and everybody's got his weaknesses." Consequently, the advice was that if you're having a rough time in one place, you'll probably find it easy going somewhere else. But as I become an experienced counselor, I came to doubt that advice, more and more because it seemed that if the guy was having trouble in mathematics, he was probably having trouble somewhere else too. When I joined the staff at ACT, I had an opportunity to examine this problem. We had reports from hundreds of colleges on academic success of students in a variety of areas, including mathematics, English, Social Studies and Science. For the first time, we had a chance to compare on a very large scale the degree to which success in these various academic areas was in fact inter-related. You can guess what the results were. There was a very high correlation among success in specific academic areas, so that if a student is having a rough time in one place, very likely he'll have a rough time everywhere else and vice versa. This suggested that when we appraise academic success by subject matter area, we are not getting really at that many different measures of a student's skill. We are getting at probably one general measure. There have been a number of factor analyses of so-called academic ability tests which pretty much yield the same sort of conclusion. A variety of academic abilities are not built into the ACT or the College Boards. At best, two factors are present. One is called verbal, the other quantitative, and they tend to be not independent of each other, but highly related to each other. Since your work is particularly in junior colleges, I want to call your attention briefly, to a third source of evidence which talks about a measure of academic success.

We had a colleague examine the prediction of success in special types of courses particularly the technical and terminal courses in junior colleges. Here is something that's different and where we may be getting away from the concentration on purely academic. He tried to examine the success in predicting achievement in technical terminal courses based upon so-called special aptitude tests such as spatial relations, mechanical ability, clerical skill, etc. as opposed to the traditional academic, meaning the ACT battery. What he found was that the two were highly different in their predictive qualities, and the traditional academic was a much stronger, a much more effective predictor of success in technical, terminal courses than were the special aptitude batteries. It was a very discouraging finding until one stops to think as I had to do in my own case that I can't put a doorknob together or a simple buzzer. I have to take the car in and tell them it's broken. I don't know anything more than that. I'm inept as far as any mechanical, clerical or woodworking skills go. And yet I took high school shop courses in these areas. I always received a grade of "B." I got a "B" because I couldn't do anything. However, they gave tests, and I could always answer tests. I could write the answers out, and I could choose the right alternative until I'd hit an "A" on the test, and the instructor had no choice. He had to give me a "B" even though I couldn't do anything. I think this is what happened in this study.

While we're talking probably about different sorts of skills, we are measuring them all by the same old ways. As it turns out the academic predictors work. People have talked for a long time about other types of achievement, and there have been some efforts via teacher ratings at trying to appraise other types of achievement. For example, faculty members have graded students on a variety of characteristics other than their grades. J.A. Davis at ETS in a larger study of this type asked the teachers to provide ratings on a

whole series of characteristics like, how industrious is the student, how cooperative is he, how independent, and how creative. First of all he discovered that such qualities could be rated with satisfactory degrees of reliability, therefore, it appears there may be a wide variety of types of accomplishment or other types of measures of student development that could be used other than that single summary measure that shows up on a transcript. However, when he took the next step, and questioned what have we accounted for in terms of individual variability that's not accounted for by academic grades, the answer was practically nothing. In other words, once he knew what grade the teacher gave the student, he knew pretty much how the teacher would rate his industriousness, his cooperativeness, and/or his creativity.

This can lend itself to several interpretations. One is that maybe we've been too hard on grades. Because grades describe a whole variety of characteristics other than the so-called academic achievement they may be pretty good measures of general worth. Alternatively, this finding may suggest that grades introduce some type of halo which will prejudice instructor perceptions and make it difficult for him to identify unique characteristics in a student. If you are an "A" student--you are a good guy; if you are a "C" student--you're ordinary. If you are an "F" student, you are a bad guy; and this general halo accounts for the type of findings reported.

This suggests the importance of teacher-set as far as students' development is concerned. Related to this are a couple of ways of looking at the Davis findings which I would like to call to your attention. Briefly, I want to review the research that was started on a National Merit Group and carried on by Holland at ACT on so-called non-academic achievements. By the time students get into the National Merit finalist category we have a varied, restricted range of talent with hundreds of students all in the upper two-tenths of one percent. Somehow, from this, we are to distinguish, make awards as to which is more excellent. This gets to be ridiculous. This partially, I think, is what prompted Holland to say since we're talking about one question maybe that separates the guy who gets the award from the guy who doesn't. Isn't there some other way we can identify differences among these students? This led him into an investigation of so-called non-academic achievements. What has the student done outside the classroom that has some sort of merit, social recognition, public recognition value? Holland identified a series of possible types of achievements of students; in science, for example, he could have won a prize in a science contest, at science fairs or something like that. He could have done experiments on his own outside the classroom. He could have adapted a piece of existing equipment to solve new problems. The series includes activities which could have taken place in the science area which describe what Holland call non-academic achievement in science because it occurs outside the classroom. It is not part of the class assignment. Similarly, there can be such achievements identified in leadership, in art, in music, in drama, etc.

In his studies, Holland developed a series of such non-academic achievement scales in approximately six areas. The question was how do those relate to measures of academic success? And the answer, as originally propounded, National Merit says, there is no relationship. The student who does well in these non-academic achievements in science, etc. is not necessarily the kid who has the high grades or the high aptitude test scores. Because of the very restricted range of talent on the basis of merit samples, one could not expect very much. That's one reason Holland was anxious to come to ACT so he could obtain good samples to take another look at this. However, when he did come to

ACT and do that across the whole range of talent, he got exactly the same answer. No relationship between these out-of-class accomplishments and the level of academic ability as measured by the ACT or a test. This suggests that there are a whole variety of other types of accomplishments, of achievements which we are not tapping at all when we get at academic achievement.

As a second follow-up to this Davis finding, partly prompted by Holland's findings here and partly in the face of some of my own experience in counseling, I determined to do a review of the literature on the whole question of what do college grades predict? Here we get into the question of adult accomplishment. Are they good measures of adult success? That of course is very difficult. What do you mean by success? How do you go about defining it? I determined to review the literature and examine whatever people have used as definitions of success. I did this in the mid- '60's, '65 or '66, and found forty-six studies that had been done since the turn of the century. Since then, I've turned up half a dozen or so others. Generally they're concerned one way or another with vocational success, oftentimes defined pretty shabbily. Sometimes, on the other hand it was done very well. Several studies of success in business were done carefully. Some of the studies of success in medicine also were done extremely well. By that I mean these studies had, I thought, a very good fix on the criterion of what is success. Medical studies, for example, had teams of experienced physicians sit in with the subjects, and with the physicians themselves. For a couple of weeks these physician teams kept track of not only their subject's bedside manner, but the appropriateness of the diagnostic tests their subjects ordered. Also, the diagnostic conclusions the subjects reached plus completeness and reasonableness of the prescriptions written were analyzed. In total these were quite thorough looks at whether the subjects were performing well as physicians or not, in the physician's sense. These studies were well done.

Some of the studies on the success as a scientific researcher, particularly in the physical sciences, were very well done. They exercised very tight controls on what is good research in the physical sciences from what is productive research in the physical sciences. I was surprised, actually, that some of these studies were so well done. Many of them center on salary or rate of promotion or something similar which I feel is a bit more nebulous, nonetheless, some of them were very well done.

Regardless of that, it did not make much difference what the criterion was. At best, we could find a minimal correlation between college grades and adult success. Generally and particularly in those studies well done, we could find no relationship between college grades and adult success. Taken together, that review and the Holland studies suggest that Davis' findings can be explained better on the basis of this halo effect. This confirms the importance of halo as opposed to an earlier conception of grades as good measures of general worth.

Let me discuss briefly a few other types of studies. These are important to recognize also. I refer you to Benjamin Bloom's unusually good developmental book called Stability and Change in Human Characteristics. Much of the work we normally have to do for ourselves has been completed for us by Bloom. He tries to trace what is known about the development of individuals. What developmental curves can be plotted for one or another type of characteristic? When does development occur? When is it rapid? What sorts of things are most changeable and when? These are unusually fine accounts, but also they are depressing in some way, particularly to educators in higher education, because he shows clearly that the bulk of the development of human beings occurs prior

to age 18. The amount of change which occurs after age 18 is relatively trivial compared to the amount of change which occurs prior to the age 18. This is true in almost all characteristics, but less true in the affective than in the cognitive spheres. More likely there might be some changes in things like values and attitudes, personality characteristics more so than in aptitudes and achievements. Nonetheless, it's a good dose of reality to ponder. How much change will occur? How much can we do? Blum helps us see that, if we learn about the human condition, we would start worrying about it long before a student comes to college. Most we would worry before he starts in kindergarten.

This introduction to studies has to do with college effects or what happens in college; what do colleges do to people. At the same there were a series of studies done some years ago on the question of attitudinal changes, as a result of college, particularly in something called authoritarianism or anti-democratic attitudes set off by the authoritarian personality studies of Sanford and his colleagues. Interestingly enough, the first studies in these areas showed what looked like very promising college effects. As a student progresses from freshmen to sophomore and on up, a decrease in authoritarian attitudes occurs, and an increase in democratic attitudes follows. These sorts of findings are very encouraging. At least until people like Telford, Plant, Lehman and colleagues at Michigan State said, "Well, I don't know whether that's college effect or not. What about student's input at college? Do they change too?" When studies were done on a total population, they discovered that the effects, associated or thought to be associated with going to college, were, in fact, general maturational effects. The same things would happen to students who did not go to college. The changes happened to about the same degree to students who went for a semester and then dropped out. These results left people very depressed, thinking perhaps a college didn't do it after all.

Similar studies initiated by Knapp and Greenbaum talked about the effects of college in producing intellectual quality. They started with the Ph.D. productivity of colleges, hypothesizing that some colleges are great for producing people who obtain Ph.D.'s; other colleges aren't so good. They examined the colleges across the country and discovered this was true. They started investigating where the Ph.D.'s--people who get Ph.D.'s--do their undergraduate work. The results indicated these colleges are not randomly distributed across the country. Rather it is a gross over-representation of some schools; a gross under-representation of others. Consequently they concluded that some schools promote intellectualism, some don't. This premise was thought to be true until Astin came along and said, "Have you really controlled all the factors?" Possibly, there might be ability differences among those schools; maybe this ought to be controlled. Or the original aspiration level of students might differ. Maybe someone wanted to pursue a Ph.D. before they enrolled in college and others did not. Why not control that or control major fields?"

When that was done, those so-called college effects washed out and no differences among colleges were found. This depressed everyone. Maybe college does not have any effect on people at all. That instigated a whole series of studies. Now, however, after ten years of defensive struggling, a few college effects have been discovered. It appears to make a little bit of difference, but not a great deal.

Some studies on diversity, studies on the nature of academic success, studies on college effects, some of which are pretty depressing, some of which are enough to make a person angry have been discussed and have implications for

your work. First, the differences among institutions--junior colleges or four year schools for that matter--the differences are so great that about the only really safe generalization is that you can't generalize. In some ways this is funny, but in many ways it is simply that. To sit and say, "Tell me what the research tells--tell me what I'm supposed to do on the basis of research," is dangerous. Chances of your school being similar to the schools upon which the research is based are not great at all. This study points out the need for every college to take a look at its own students and its own programs and to draw conclusions accordingly.

Second, these studies emphasize the importance of input and the importance of what characteristics the student body possesses. In general, the developmental trends that can be projected are the developmental trends that are apparent at the time the student enters college. The characteristics he has been developing and that have been promoted, that do look like they're special characteristics, are special. These usually turn out to be those characteristics most likely to show further development through college. Those that have not been coming along, that he has not shown strength in, are not likely to be involved very much at all.

Third, the study to stress the independence of academic achievement from other types of measures of success is a very important one. What one knows and what one does with his knowledge are probably two different things. It is about time that we wake up in education to the fact that when we put a grade on the transcript, we have done a pretty poor job of describing what has happened to a student, both to him and to society.

Fourth, this being the case, it seems not only desirable but essential that we now begin to broaden our base of evaluation in higher education and start looking at types of student development other than the purely academic type of development. As we do that we thereby broaden the variety of reinforcements we're allowed to supply the students, the variety of ways in which we can encourage students, the variety of ways in which we can promote learning through reinforcement.

Finally, there are implications in this for what we call common educational objectives, a series of objectives we expect every student in college to meet. If we have such objectives, and most colleges pretend to currently, with what we know about developmental psychology now, we had better be very careful about stating those. When these objectives are stated, they must be at a low level. Either that or not have any at all. An important illustration of this is in Lenning's ACT study on the question of English skills development. Almost every college in the country says that all Freshmen take two semesters of English, or three quarters or a full year of English. That's a requirement for everybody or one of those common objectives. Lenning was interested in how successful we were at the end of the year in promoting that type of development. The ACT English test was used as the measure. You have the measure beforehand; you take it after a student has finished his two terms of English. It is as good as any of the other measures of English achievement. Nothing is perfect, but it is as good. It is as good as a theme or any other type of English test.

The question is then, when Lenning did this at a variety of colleges, there is significant gain. This I would not want to dispute, students do change. I'd have to ask the question how much. What are we really striving for to

best illustrate: suppose we set an absolute level of achievement in English as being desirable--across the board. Therefore, if we cut right in the middle at the mean or median of the level already achieved by the entering Freshman class, on ACT English scores--half of our students are already there. Half of the students meet our objectives as far as the level of proficiency we expect. Now is it possible to bring the other half up to that beginning point and call that minimal achievement? Lenning said he showed that level was not achieved, and, in fact, if the rate of growth were consistent from year to year, it would take ten years of English training to bring the lowest ten percent up to that level, or up to where on the average, the entering freshmen began. So to have common educational objectives, one has to set the level of the low.

Second, when we look at this variety of studies, we have to think about some educational folklore and educational commitments. This may be more true of four-year colleges, but it definitely relates to junior colleges too. We talk about the educational mission being the development of well-rounded citizens. However as we look at more and more literature, and particularly at the independence of various types of measures of achievement, we wonder about the **rate** of development and developmental histories of people. We wonder whether a man's nature is to become well-rounded, or whether he's to become somewhat lopsided. The way I read the literature, man has a proclivity to become more lopsided rather than well-rounded. When put that way, it does not sound very good. Psychologists have ways of putting it to make it sound better. They talk about people searching for their own identities, or finding ways in which they are unique. Each of us can identify somewhat with that. We don't want to be like everyone else. We want to have some things which make us different from anyone else. Do not make me like everyone else; give me something that makes me different--to stand out, to be myself, to be unique. The research literature gives us this type of perspective on what the nature of our mission may be.

The final implication that I see from this is that the educational payoff, the place where we're going to profit, will come much more if we take as our mission the concentration on developing individual strengths rather than the overcoming deficiencies. Too frequently, I think, in the past, the focus has been on what is the student weak in and what can we do to make him stronger in that. All the way down the line we fight his developmental history, when we do things like that. We are hurting ourselves thereby. We are more likely to expose him to a series of more failures, things in which he has not done well in the past. If we could choose the things in which he has achieved and attempt to exploit them, I think the chances of our having a positive impact on him are very much greater.

This perspective on the concern about the role and sorts of measurement with which we work has been concentrated largely in the academic area. It has produced the sterotyped educational processes, and in some ways sterotyped people, and in many ways valid charges of educational developments. To further discover what students are like, now is the time for us to branch out into other types of characteristics.

What do we need? What do we need in the way of other types of measures? Hopefully fifty years from now, this topic will still be incomplete. There is no way in which it will be resolved once and for all. When one gets final answers in anything, I get very worried.

Some of my best thoughts on this topic are taken out of the context of intellectualism and have come from the humanities. At one time I enjoyed a novel by C. S. Lewis called Out of the Silent Planet. In the story a man leaves earth and comes to another planet where he meets with strange beings. These tall, white, thin, wispy beings are called Sorenai. Sorenai are people who are thinkers. They theorize; they conceptualize; they intellectualize; they are really top notch in the way of thinking. However, the character from the novel lives with another group, called the Harasa. Harasa are, visualized by me as hippopotamus-like except they have extra feet, little gadgets running out here and there from their bodies, and no noses. They are the artists, the poets, the cultural expressionists of the society. They paint, they philosophize, but when it comes to an intellectual explanation, they always say "you'd better go ask some Sorenai about that. You know I'm expressing a feeling here." The third element of the society, the Flitry, are seen like woodchucks--very busy. They have several arms and legs, and they are around hammering and fixing and building and doing all the time. These three diverse elements lived together very well in this happy society which he visited. However they are quite different in their orientations and this left an interesting impact on me. This is what we must do when we think about what are human characteristics. I think Lewis was pretty close to them when he defined the Sorenai and the Harasa and the Flitry. Lewis presents characteristics dealing with knowledge, understanding, reasoning, cognitive or thinking measures which are an important part of our society. Likewise other characteristics of affective or feeling measures which remind one of the Harasa e.g. the attitudes, beliefs, feelings of the people, are shown to be relevant and important. Then we have the accomplishments of the Flitry's, the motor or doing type measures which seem to be important. Consequently, Lewis was saying there are three types of beasts on the planets--thinking, feeling and doing. Probably Lewis is talking about the Silent Planet right here on earth where we find these characteristics of thinking, feeling and doing.

Of course, our traditional measures concentrate on only one of those. As far as the psychometrics, and mathematics and other elegant treatises, which are available on how to measure educational achievement in a traditional sense, go, we have done a good job. Ebel's book called Measuring Educational Achievement is probably the best, but there are a number of them with pretty straight rules. These are basically good but with limitations. Our concern is to be about things that go on beyond these. I want to discuss appraisal of the individual, with concentration more on the feeling and the doing parts than on the thinking or cognitive parts. In counseling much of this appraisal is done in your individual interviews versus a more formal objective approach to appraisal. In counseling you need enormous flexibility in the appraisal process, simply because in the situations you encounter, people are so different. Through the interview what you want to ask in order to make some sort of appraisal depends very much on what that person is like, what situations he's encountered, why he's there, and what his situation has been. Therefore, in the appraisal process flexibility in the counseling interview is highly important. However, the validity with which you can interpret this type of material, the meaning of what is communicated to you, is increased greatly if you have some sort of normative information from which to proceed. Without a great deal of experience, it is difficult to discern when a report is unusual, when a person had told you he has many characteristics, what looks like a really important talent and what looks like an ordinary sort of indicator. Without some sort of normative background, it is difficult to interpret

what you could pick up from the individual appraisal interview. We cannot be satisfied with the interview as the only appraisal device, but neither can we afford to ignore the fact that a great deal of the practical assessment which occurs will have to be done this way. This points out, where we need to go with our more formal appraisal devices. Until we become aware of how we can do this in a more objective way, we will have to use the less satisfactory, informal appraisal devices from the interview. Regardless of which method, the interview or the formal appraisal method, is employed, the principles which will lead to satisfactory judgements of appraisal are pretty much the same.

Two principles of appraisal are noteworthy at this point. The first one says if you want to know something about a person, ask him. In 1954 my first Master's candidate wanted to work on a measurement problem. He wanted to develop a scale for identifying or predicting alcoholism. He had two populations of alcoholics who had responded to the Minnesota Multi-Phasic Personality Inventory. This inventory contains 566 true-false questions, for example, is your sex life satisfactory? and have you stopped beating your wife? He had a few alcoholics answer true or false to all those questions. Then he had a good sample of normals--non-alcoholics who answered those questions. Following the right principles for developing a test for alcoholism, we set out to determine what percent of the alcoholics in this first sample said true to each question. Then what percent of normals said true to each question. Statistically we could compare the responses and where a difference showed a higher percentage of the alcoholics said true or a lower percentage, that item could be retained as an item to put on the alcoholic scale. My candidate went through that whole process and picked out about sixty items where he got significant differentiation. He could have called that the alcoholic scale, and then he could cross-validate it on this other sample for alcoholics. He could score them on those sixty items to see what percent he could hit if he'd cut it at a certain point. He had good success with cross-validation and hit about 79 percent. Therefore, it appeared it was working out the way it was supposed to do to fit the scale. The question arose about what shall we include in the way of tables to show this. I felt the individual items should be included. This would show the differentiation on the percent of alcoholics and normals who responded to each item. The individual items from the inventory could be an aid to someone using the study in the future. Up to this time the items were simply numbered items. The items had not been examined. The items were written out, and the percentages recorded. It turned out that there was one item more accurate than the entire scale. One item hit 81 percent as opposed to 79 percent. This item was, "I've used alcohol excessively in the past." At this time I formulated my first principle that if you want to know something about a person, ask him. If you ask him "are you an alcoholic?," the chances are you'll find out as much as if you asked him how his sex life is.

The other principle, an important one, stated simply, is that the best predictor of future behavior is past behavior which most closely resembles it. Very sound principle and can be dressed up in fancy language that follows the rules of developmental psychology. It states a person's development isn't chance. The history of an individual has meaning; it is biographical psychology at its best. It is a look at what his life style has been; it is going to predict what his future is going to be like. In a very practical way all you have to say is the best prediction for future behavior is past behavior which resembles the future behavior. You can see it experimentally. For

example, any time you want to predict college grades, what's the best predictor? High school grades--always. If you want to predict graduate grades, use college grades. It's a similar sort of process. If you're stuck with aptitude tests to predict grades, what type do it best? So-called reading comprehension tests where you happen to read a paragraph and answer questions about that. It's just that it resembles very closely the situation in college where they had to answer questions about what they had read.

If you take those two principles, you can go a long way toward the type of needed individual appraisal. These principles mean that we need to be aware of, and to determine if necessary, the environment where for a given individual learning experiences have most likely occurred. Where the characteristics that are of interest to us might be displayed. We need to know what that environment has been or we won't be able to ask the right sorts of questions. We won't be able to depict the sorts of experiences which the individual has had. We'll need to know, in particular, something about the sort of schooling he had had, the neighborhood he has been in, work experiences he has had, the peer group with which he has been associated. With kids is where the action is. In schools, in neighborhoods, at work, with their families, at their recreation, and in their peer groups is where they build their developmental histories. All of the important individual characteristics are suggested by principle number two, also you need to know something about an individual's developmental history and you need to know that it occurs in these contexts. Unless you know what an individual's developmental history is you won't know what you might be able to ask him, what would be useful in helping him. How was his time spent in these environments, which of the activities dominated his time in those environments, which activities were chosen by him, which activities were imposed, are also important. Activities which were imposed have a little different meaning from these which were chosen. Finally you have to see what was accomplished, what sorts of things in these environments were well done, which ones were done poorly and with what effect. How did he feel about that. For your understanding of what the individual is like, this is important. Eventually, then we need to ask what themes developed in our review of this developmental history which defined his own uniqueness, his own individuality. This is what these principles mean. However, it is more complex that I've made it sound. However, a few examples of how these principles can be put to some practical utility or have been put to some practical utility should help dispel some of the complexities.

Self-appraisal: based on Principle Number One. You want to know something, ask the student. We wanted to know what students needed some remedial assistance in educational skills. We have a clinic and I was able to institute an experiment wherein we tried to test the validity of that principle by giving this as our diagnostic measure to half of our entering freshmen. To the other half we administered the usual four-hour diagnostic battery including reading tests, tests in English mechanics, a theme, and a mathematics test. In the study the educational skills people did not know on what basis the student was selected for the clinic. They knew why he was there, for reading, or mathematics or what have you, but they did not know on what basis. They were not informed whether the student came because the test results said he ought to go there, or because the student himself said he ought to go there. At the end of the term, the personnel in the clinic put each student in one of three categories: 1) Yes, he really belonged here; 2) I'm not sure; and 3) No, I don't think he did belong here. We were interested in comparing, taking that as a criterion. They had the students for the whole semester, did we make a good judgment in their showing up in the clinic? In the reading

area, on those that were assigned on the basis of the reading test, 74 percent were said to belong there, 20 percent question mark, 6 percent, no. 74-20-6 on reading. Using the self-rating: 91-9-0. English Mechanics, the test 76-24-0; the self rating; 94-6-0. Using the English theme: 65-21-14; self rating; 83-10-7. Just to be perfectly on the mathematics test: 65-19-16; self-rating: 61-20-19. Math's a tough area. Maybe we used the wrong questions, but it seems to be a very difficult one in which to work. The test did not do very well, but neither did the self-rating. We are not satisfied with this area. This is an example of how using student self-appraisal can be helpful in diagnosing educational deficiences.

SEE TABLE ONE: EXAMPLE TWO - IDENTIFYING COMPETENCIES

On this table there are options to determine whether students have some special competency. This is opposed to a kind of a hidden talent search in that it says do not look for hidden talents. Maybe at age two or three you can look for hidden talents; however, at age 18 the talent inherent has been discovered probably through natural processes. So we ask simply what do you do, what do you do that would be suggestive of certain types of important talents. The ones illustrate what could be done. These are labeled electrical, mechanical, woods and metal, secretarial and clerical. You find out by statistical processes in part whether you've got the right areas or not. Does one thing go with another? Does this seem to make a homogeneous scale one way or another? Sometimes you are forced to know whether those are good areas or whether you need some other area if these turn out to be pretty fair. Therefore, if you are interested in electrical-mechanical types of competencies, how could a seventeen or eighteen year old demonstrate what is reasonable for him to have had some exposure to and give him that as a question. This is the building of a kind of formalized biographical history. It's a technique that can be widely applied to the assessment of individual characteristics. Showing what the developmental pattern for a student has been is very important and useful if we learn how to interpret them in terms of helping a student plan for his future.

Essentially each individual goes through a whole series of experiences, and it is impossible for a person to develop every potential that might be there. Because there are only 24 hours in the day and only so many years allotted to everybody, some choices must be made. Choices are made every day. We do have, then some sort of developmental history which reflects the sorts of choices we have made. Why do we make them? For many years as counselors and psychologists we have talked about the fact that we have interests, aptitudes, talents and values. Because of these we make decisions which have some personal meaning to us. Frequently we do something well and we like to do things we do well, or we find something personally rewarding, or something that we are interested in, or we think is important to society and thus important to us, are all basis for our decisions. As a result of our history, we develop something called "our bag." You have your bag; I have my bag. Each has his own bag. It might be wise for us to search out exactly what that bag is and to see everything we can do to promote his identification with the full understanding, of course, that once you do that you are shutting off some other opportunities for the individual.

TABLE ONE

Example Two

Identifying Competencies

Questions 89 to 148 describe skills or competencies. Indicate whether or not you can perform the activity named by using the following code:

Yes, this is true for me.....1
No, this is not true for me.....2

Electrical-Mechanical

89. I know how to use a voltmeter.

90. I know how to use a micrometer.

91. I can adjust a carburetor.

92. I have changed a part on a car engine such as a water pump or a generator.

93. I can "tune-up" a car.

94. I can build a radio.

95. I can make simple electrical repairs.

96. I can do simple home wiring.

97. I can make simple plumbing repairs.

98. I can use welding equipment.

Wood and Metal Work

99. I have used wood shop power tools such as power saw or lathe.

100. I have operated metal shop power tools such as a drill press.

101. I can refinish varnished or stained furniture or woodwork.

102. I can read blueprints.

103. I can repair furniture.

104. I can make mechanical drawings.

105. I can form sheet metal objects using the proper tools.

106. I can operate a metal lathe.

107. I can build a major piece of furniture (bed, dresser, etc.).

108. I can repaint a car.

Secretarial and Clerical

109. I can type 50 words a minute or more.

110. I have operated office machines such as an adding machine or a duplicating machine.

111. I can take shorthand.

112. I have operated a cash register.

113. I have filed correspondence and other papers as part of either regular or voluntary job.

114. I have held a job in an office.

115. I can operate simple data processing equipment such as key punch, sorter or collator.

116. I can operate a bookkeeping machine.

117. I can operate a calculating machine.

118. I can operate a dictaphone.

DISCUSSION

Question: On what basis did you select the items a typical 18 year old would have experienced?

Dr. Hoyt: We went to high schools in our immediate area in order to do this. We talked with teachers, particularly in Shop and Woodworking, asking what would kids be able to experience. We talked to people at the vocational-technical schools in the area about what they would expect people to be able to have. Also we talked to some two to three hundred students, as to what types of experiences they had had. On that basis, the choices were made. You have to make the use of the best consultants available realizing you are going to miss sometimes. Where you have made mistakes in your judgments or your advisors were not very good advisors becomes apparent after you have tried these things out for a while, and then you have to modify. If you end up with an item where one percent says yes I can do that, 99 percent says no, you do not have a very good item. Maybe it differentiates the very, very talented from the others, but it doesn't work at all in 99 percent of your people. In this case you would look for something that might be a little more typical. However you have to seek out people who might know what is reasonable. That is why I said you have to know what the environments are of the people to whom you are interested in applying your measures. These measures are built on what opportunities students could have.

SEE TABLE TWO: EXAMPLE FOUR - IDENTIFYING NOTABLE TALENTS

Question: With Table Two, are you using social service talents as a predictor whether or not someone would make a good counselor?

Dr. Hoyt: These again, are only illustrative. I included social participation and social services particularly to suggest that you need not be stuck with English achievement or mathematics. You can talk about things that have relevance. Personally, social service is not a good one if you talk about the field of counseling. You can question what has a person done to show that he has the interests, committments or abilities that make a person successful in working with someone else for the client's benefit. Does he have a history which supports this type of activity, and, if he doesn't, my guess is he may have drifted into this area or he may be in it for some reason other than a natural support for his own developmental life style. I have not tested for this yet. I will be very anxious to get a test on it though. This will work better in selecting counselors than a GRE.

Question: Also, these might imply manipulation of other human beings, or that this person has had a need to manipulate human beings all his life; therefore, he would be a very poor choice as counselor.

Dr. Hoyt: Yes, it could. It would be well important to test. This is the only way you can settle this problem because this could lead to a semantical struggle. Is it manipulation or is it helping? Is there actually a difference

TABLE TWO - Example Four - Identifying Notable Talents

Questions 100-199 also deal with experiences you may have had in college. Read the items in each list. Then indicate which ones are true of you by blackening the appropriate space or spaces on your answer sheet.

List 2. Social Participation

110. Actively campaigned to elect another student to a campus office.
111. Organized a college political group or campaign.
112. Worked actively inan off-campus political campaign.
113. Worked actively in a student movement to change institutional rules, procedures, or policies.
114. Initiated or organized a student movement to change institutional rules, procedures, or policies.
115. Participated in a student political group (Young Democrats, Young Republicans, etc.)
116. Participated in one or more demonstrations for some political or social goal, such as civil rights, free speech for students, states' rights, etc.
117. Wrote a "letter to the editor" regarding a social or civic problem
118. Wrote a letter to a state legislator or US. representative or or senator about pending or proposed legislation.
119. Worked actively in a special study group (other than a class assignment) for the investigation of a social or political issue.

List 4. Social Service

130. Worked actively in a student service group or organization.
131. Worked actively on a charity
132. Worked as a volunteer aide in a hospital, clinic, or home.
133. Served as a big brother (sister) or advisor to one or more foreign students.
134. Organized a student service group.
135. Worked actively in an off-campus service group or organization.
136. Worked on campus or civic improvement projec
137. Participated in a program to assist children or adults who were handicapped mentally, physically, or economically.
138. Voluntarily tutored a fellow student.
139. Received an award of recognition for any kind of campus or community service.

List 5. Scientific

140. Built scientific equipment (laboratory apparatus, a computer, etc.) on my own (not as part of a course).
141. Was appointed a teaching or research assistant in a scientific field.
142. Received a prize or award for a scientific paper or project.
143. Gave an original paper at a convention or meeting sponsored by a scientific society or association.
144. On my own (not as part of a course), carried out or repeated one or more scientific experiments, recorded scientific observations of things or events in the natural setting, or assembled and maintained a collection of scientific specimens.
145. Author or co-author of scientific or scholarly paper published (or in press) in a scientific journal.
146. Invented a patentable device.
147. Member of a student honorary scientific society.
148. Entered a scientific competition of any kind.
149. Wrote an unpublished scientific paper (not a course assignment).

between manipulation and helping? The question of who helps the client is the important issue. We can look at that without necessarily looking at the counselor. We can look at a client, and we can ask in what ways has he changed as a result of interaction with the person. I would like to compare people who have had this type of experience with those that have not without worrying about what that interaction has been or whether the counselor is manipulative.

A person is concerned about people, or if he wants to be of service to people, if he feels he has a role along this line, this will show up in his history somehow. If a person's life style serves his needs, and you can start identifying these early in life, they will follow a relatively consistent pattern.

GROUP REPORTS ON WRITING MEASURES

GROUP ONE:

Our problem area was one of flexibility and adaptability to different kinds of situations. This is tied together with problems close to the colleges in the different situation which students face when they are lifted out of one environment and placed in another, or from one cultural group into another. The skills for survival and success and good adaptation were the focus of our discussion with special reference to the ability of the young person from rural area to be able to make his satisfactory adjustment in an urban one, and the ability for some instrument to predict success and degree of success. Many questions ranging from do you enjoy walking through the woods, to liking crowds, styles, waiting in line, does waiting in lines bother you, were listed. Do you like to get in a car and go where you want to go? Doyou like to be in the situation of people rushing to and fro? The last item we have is do you simply like a beer at night because looking perceptively at the situation of the 18-20 year old in the city as against the suburbs, it was brought out that beer isn't any longer a typical mode.

Dr. Hoyt: It sounds like a good characteristic, and I'm sure one with meaning. The process you are going through is typical when you try to develop a scale of this sort. You listen to each other say, "Well, I think it is this" and you say, "what are you talking about, it's not that at all" and that does help. In the long run a person gets further along on this type the more he can tie his conceptualization of the concept to some specific type of behavior, or a reflection on the historical past. What is the extent to which this person has sought out change as opposed to security and sameness?

I am not sure offhand what particular behaviorial samples I might pick out to explore that. Surely change is involved in getting outside one's neighborhood. Therefore one might ask, "Have you visited six or more states?" This certainly would be a way to get started. In general we have found that as you ask for "what do you think-type-things, or how do you feel-type-things," you have to ask more questions. These questions become more dubious in meaning. However, if you say, have you done something; if you ask the behaviorial items, usually eight or ten suffice. If you went on a strictly how do you feel type of instrument forty or fifty items would be required to be reliable. A person can say yes or no about having done a certain thing, but he may have to scratch his head and come up with several different answers when you ask him how he felt about it.

To accomplish this, one would have to ask about the kinds of family environments the kids are from and how they could demonstrate adaptability in their families. What has happened to them when the family had a plan in mind and the student had a

plan in mind. Was it easy for the student to adjust? Did the student find this always a disappointment? What are peer group interactions or adaptability to the peer group? This might be demonstrated by the conflict of personal expectations of the group. Recreational opportunities, neighborhood confinements, represent other typical behavioral data in which students engage. Give students a chance to demonstrate flexibility or rigidity.

I think, to develop items of this nature it is best to reflect on the types of opportunities which say behaviorally, what would you (the student) have to face. Then the problem is how can these be put into terms to which one can answer yes or no. I behave flexibly, or I have not been able to behave flexibly. I feel this is a way of getting at this problem and an additional way in which these items may be strengthened.

At one time we started to think about a measure to determine how a young person can make it, or the extent to which he can demonstrate the skills of being able to win the game. We kept in mind a historical antecedent and developed the theory that activity or behavior from the past has a tendency to be repeated in the future. We tried to develop a scale of interpersonal competence and we came up with a number of items. We were not quite sure of the phraseologies of the time. We developed six pages full of them in incoherent fashion.

We got into trouble because we pinned them to behaviors. For example, we decided to look at interpersonal competence. Our historical antecedent was to share a room with a sibling has a way of demonstrating interpersonal competence in the past. Next step develop a question. For example, I have shared a room in preference to living along when I had a choice. Co-operative versus individual sports. I prefer (not prefer) the team sports. I have played basketball better than track. At this point we are getting at recreational groups. We developed items like: I enjoy listening to all kinds of people. I enjoy talking to all kinds of people. I enjoy being subjective.

We made items more direct and stronger by saying, I seek out people with markedly different background. I am able to take criticism to alter my ego. I am able to take feedback, or I can give feedback constructively in a group situation. These also have to be related to activity or behavior in the past. In line with this, we developed such items as: I can usually engage others in conversation. I do usually engage others in conversation.

Because these items are tricky to write, you will spend much time developing them. Phraseology has to be right. You have to describe a situation to primarily strangers. One item, if you were in an elevator with a strange person, would you converse with him or not, must be worded have you, as opposed to, would you. Instead of saying I find it easy to categorize people, or I do categorize people, we must state it more behaviorally. A possibility might be: When I am in a situation, I try to size people up particularly into categories. Now our description is more behavioral.

Other items which might be included in such a battery might be as follows. 1) I have dated more than ten members of the opposite sex. 2) I have won some sort of most popular award. 3) I have been asked by friends to help them resolve an interpersonal conflict with someone else. 4) I have been invited to social events of interest to me. 5) I am sought out by others as an advisor.

Question: What about when asked to give my advice on a problem, I have been able to accept opposing views and let them do their own thing even though I would not have chosen those alternatives. Or, in the past in the solution of common problems, I have been willing to go along with the group.

Dr. Hoyt: You are almost forced into something like that, but write your item more behaviorally. However trouble develops trying to do this also. To illustrate take a negative item like, "I can think of six or more people whom I go out of my way to avoid." This item is getting at the same kind of thing, and is tied down more explicitly. Building such scales is really not simple; it is a complex task. It is particularly important to get participation from some representatives of the group to whom you hope the scale will eventually apply. Hopefully they will have the perspective and background you desire. Oftentimes they can give you clues as to directions you really want to go.

Question: Is it a problem then that we need to decide on a common person we would want to measure?

Dr. Hoyt: Yes, because the characteristics themselves are demonstrated in different sorts of settings that probably will give you trouble. When you go this route, you must capture the settings where the exercise can take place. However, it is unlikely you could get a really satisfactory coverage for really diverse groups.

Question: Then the kind of skill you are measuring has a lot to do with it also? It is easier to devise a scale for electrical or mechanical skills like, "I can repair a carburetor," because that is a concrete task with definite relationships. In dealing with interpersonal conflicts everyone has some idea what's involved, but recalling specific past performances which reveal it is difficult.

Dr. Hoyt: I think so, and, particularly, if you are trying to think of very diverse groups. The developmental tasks which are available for 18 to 20 year olds are really quite different from ones the 40 to 45 year old person would have encountered; therefore, the signs you would pick up for a strong point or a weak point will differ quite a bit for just the age groups.

Question: I can't think of a past behavior which deals with how an individual relates to authority figures or also something about personal defense structure. Is an important part of interpersonal relationships the ability to take criticism?

Dr. Hoyt: Except indirectly, the interpersonal competency part comes not from the way criticism is given or taken. You have to remember what it was you were after. It is the outcome at the end. Did you end up so that your friends allow you to talk? Do you have communication still open in spite of the fact there were conflicting views? I would sure give constructive criticism somehow. I don't have any friends.

You go through this type of enterprise where you play out what looks to you and your advisors as the best way of going about it. A series of items you think are pretty good. You may not get full agreement, but you get reasonable agreement about the way to say it. Then this is edited properly so you have said about the best you can say. You have to revise some after that to make sure you have what you think you have. If you aren't way off, maybe someone says, "I don't think that's a good item, but, by and large, I will buy the other items." That's a good thing. You are okay and ready to go. Then you try it out and in the absence of anything else, you use this total score as a criterion, if, in fact, these pool judgments are right and there is a generally good group of items.

If you set up your scale on a fourfold basis (Yes/No for this item and above average/below average for the group as a whole) and you find that items aren't going together or an item splits the wrong way with more of the lows saying Yes than the highs saying Yes, this is statistical evidence of a bad item. This item doesn't relate to what the rest of the items are getting at, and your judgment indicates the rest of the items are getting at universal results. You throw this item out on that basis. This is the beginning step and as far as probably one needs to go at this stage. Eventually more predictive levels are reached, but that requires an internal validity step.

Whenever you are calling for self-report the question of distortion is something of concern. A surprising number of looks have been taken at the concern. Particularly, do students constantly distort when their values are saying, "Yes, I've achieved that," or "No, I have not"? The evidence is clear that students usually tell it like it is. Rarely are there conscious distortions. There are errors reporting and other inaccuracies. In one of the best studies of this kind as a part of a scholarship selection program students were asked for their special accomplishments. Students have considerable motivation to distort. (Because they'll get a scholarship dependent on the way they answer these questions.) In those instances the degree of distortion has been minimal, practically none. There has been almost as much distortion in the negative sense, that is, not reporting something that did happen, as there has been in a positive sense of reporting something that did. And by and large no distortion in either direction. Students will tell it like it is. You do not have to worry too much about their trying to snow you or their giving you a bad word. I hate to generalize that too far. It has not been tested in all cultures and under all conditions. As yet I have not seen any evidence of places where students consciously try to give a bad picture.

This is a tough task, and I am reluctant to leave it. I want you to experience what it is really like to nail down some of those items. The key to writing this type of item is putting it into a behavioral trend so if the person does feel that way about it, he will have demonstrated some sort of behavior in our appropriate setting. Writing these items does take some experience.

A few items on the persuasiveness scale which might clarify the types of tasks I was suggesting to you might include:

1) I convinced a teacher or principal to change a decision they had make about me.

2) I convinced a group to adopt one of my ideas.

3) I have directly been involved in bringing about a change in school rules or regulations.

4) I have persuaded friends to join me in an activity in which they had expressed disinterest.

5) I might have won special recognition for selling tickets to a school concert or production.

6) I have convinced my parents to change their decision about me or a decision they made about my brother or sister, or a decision they made about a vacation or family outing.

7) I have recruited volunteer workers to help in a neighborhood or community project.

In the area of program evaluation there are a few rules and regulations which will aid understanding. Some examples might show what this really involves. Always in program evaluation the beginning point is the general concept, that is, your problem, what it was you were trying to get at,--the general nature of your concern. You start with whatever it is. It could be effectiveness of counseling, teacher effectiveness, effectiveness of a curriculum, etc. Your first step is try to identify, as best you can, the elements present in this concept.

Then as a second step, insofar as possible, identify their dimensions. Both are difficult steps, but they go together as a way of trying to pre-conceptualize the problem you have in mind. You accomplish those steps by using whatever means-fair or foul, but usually they are accomplished through theoretical considerations of one sort or another. Theoretical considerations can mean the way it is supposed to be according to the books or according to your daydreams. What practical experience or cases you have had. You consult with other people to find what they have had to say. Your own personal daydream as to what you think that concept or problem is all about. Use any of those sources,--fair or foul. They will give you some way of getting started on what the problem is, of identifying what the elements are, and what the dimensions of those elements are. Of course, doing it that way means you are never going to be right because there is never going to be a really right answer to possibilities for conceptualization, for really drawing out, for understanding. Total problem or concept possibilities are endless and never exhausted, but it is necessary to get started somewhere. Do not get stuck on the idea that, because it is incomplete, we had not better do it. At this point nothing ever gets done. Somehow satisfy yourself that this is not the whole answer, but, may be in large part, a delusion. It is a place to get started at least. Then one can discover what ways he has been deluded, and the next time around make it better.

The next step is to determine who can give you some answers. Who is qualified to respond. Who is qualified to give information. What is the probability that this source can give you an informed response or a dependable answer.

For step three: bearing in mind who can give the answers along with what elements and dimensions you are after, develop some pool items. The development of the pool of items is again messy; there are not definitive answers. How do you do it? Review the literature insofar as people have written about or have had thoughts about developing items. Consult the experts or the best experts you can find--maybe your nextdoor neighbor. Do the best you can. Become personally acquainted with the operation. If the operation is something about which you do not know very much, spend a few days surrounded in the milieu or in the problem area for which you are to write items. Frequently, if you are trying to develop a pool of items, start with a simple open-ended questionnaire where the respondents you are worried about give openended responses. This can be used as a source for developing a pool of items for eventual use in your instrument.

The final step is to test the developed items. A simple test can be a review by colleagues. They will comment on a bad item, a good item or an item not understood clearly. In a small target population, try out your items for a few things. Ambiguity--do people understand what you are talking about. Do they have questions about what you meant by that item? Try out items for offensiveness. Do

respondents object to answering questions like that? Does it bother them? Did they feel the item caused undue invasion of privacy? Try them out to see how much time responding takes. In addition to tryouts of that sort, there are certain statistical tests which ought to be done with these items. Some statistical test to know if the item belongs in the scale and whether it works or not, whether it does the job. In general on that type of test, search out professional statistical advice.

Some principles for writing items include the following:

First USE SIMPLE BUT CORRECT LANGUAGE. Do not insult your reader, but make his job as easy as possible. Do not water your language down so much the reader thinks "This is Dick and Jane." This is especially true if your population can read beyond Dick and Jane. However, do not overwhelm them with your fancy vocabulary, the profundity of your own philosophical inclination.

Second AVOID FUZZY TERMS. Terms like often, sometimes, perhaps mean different things to different people and should be avoided where possible.

Third AVOID WIDE DOUBLE-BARRELLED QUESTIONS. Many times a question will elicit two responses; only one of which may be of interest to you. Last year I wrote an item which is a very good example of this. The university governing bodies are too dominated by a few professors who have inadequate understandings of their students. That item elicits two responses. "Too few professors are covering university governance;" and "Who have inadequate understandings of students," which leads to how professors do not understand students. Two responses are possible to one question. That is a double-barrelled question. If both aspects are important, ask two questions. Otherwise stick with the one you want; either "Too few professors" or "Professors having inadequate understanding of students."

Fourth AVOID PREJUDICING THE RESPONSE OF THE SAMPLE. This can occur in a number of ways. You can ask questions like: Do you think improvements can be made? Yes or No. No matter what, the answer is always yes on something like that. The answer tells you nothing except there is not anything which exists that cannot be improved. People know that, and they will tell you that. However, a more subtle type of prejudicing occurs when we establish sets by the way we arrange our questionnaires. All questions are framed positively; therefore, questions are answered YES if good and NO if bad. Try to avoid this by alternating the way questions are written; some negative, some positive. However, for the sake of alternating positive and negative, do not distort your questions or make them awkward. If it turns out you have to write more true questions than false questions, do not panic, but insofar as possible try to avoid that situation.

Fifth ASK YOUR QUESTIONS IN AS SPECIFIC A LANGUAGE AS POSSIBLE. An item from a teacher evaluation was, "His speech needs improving." It may be a perfectly good item, but what does it tell you? It doesn't tell you anywhere near as much as one of the following items: He lectures in a rambling fashion. He lectures in a low monotone. Both of these are indications his speech needs improving and tell you what needs improving and what the problem really is. Try to be as specific as possible. Avoid using generalizations.

Sixth DON'T ASK FOR FINER JUDGMENTS THAN CAN BE MADE. Studies suggest it is very difficult for human beings to make judgments on a continuum of more than five to seven points. At maximum use seven points, but five may be closer to what the respondent can really do. This is in relation to appraisal for use with the

individual to identify some characteristics useful in your counseling with that individual. There is some advantage to using a 5 to 7 point scale. You will have increased reliability over what you would get if you had a two point--Yes/No or True/False. Some of the behavioral items indicating have you or haven't you, would be a little more reliable if your scale were never, once or twice, three or four. A 5 to 7 point scale adds to the reliability of your device. Another thing that adds to the reliability is permitting "no opportunity to observe" or "not qualified to judge" type response.

Here are a few examples of how these rules have been applied. The general concept or problem is to evaluate instruction for the purposes of improving it.

Step One as you recall, was: What are the elements in evaluation? The goals in education are said to be of two types: cognitive and affective. The elements, therefore, about which we are concerned are cognitive development and affective development.

Step Two. What are their dimensions? Again search the literature, discuss with colleagues, teachers to whom this instrument will be applied. The dimensions chosen will differ from one place to another depending on what is stressed. With these dimensions what was determined in the cognitive domain?

To illustrate note Table Three below, covering questions 59-66 from the instrument identified.

TABLE THREE - Student Reaction to Instruction and Courses: Items 59-66
(Each item to be answered True or False)

59. Gaining factual knowledge (Terminology, classification, methods, trends)

60. Learning fundamental principles, generalizations or theories.

61. Learning to apply principles to solve practical problems.

62. Understanding myself--my interests, talents, values, etc.

63. Learning attitudes and behavior characteristic of professionals in the field most closely related to this course.

64. Developing skill in effective communication.

65. Discovering the implications of the course material for my personal and professional conduct.

66. Gaining a broader understanding and appreciation of intellectual-cultural matters. (Music, science, literature, etc.)

From this table one can find in the cognitive domain: 59--factual knowledge; 60--principles and theories; 61--applications; 64--communication skills. And the affective items can be identified as: 62--self-understanding; 63--professional attitudes and behavior; 65--implications for personal professional growth.

One item, No. 66 was a mixture of both cognitive and affective. This item shows a broader understanding and appreciation. This illustrates the dimensions of the two elements we have identified in this problem.

Step Three. Who can give the answers we need is the next thing to decide. We must question what do we need? We want to evaluate instruction. How do you do that? You have to answer the question: Were the objectives of instruction met? To the degree that the objectives of instruction were satisfied, we say instruction was successful. To the degree that the objectives were not satisfied, we say instruction was not successful. Now we are ready to answer the question, "who can give us answers about what the objectives of instruction are?" This answer comes from the faculty member who says what his course is for, what his objectives are, what he is trying to do. He sets the objectives. If you want to know what are the objectives, ask the faculty. If you want to know were the objectives dealing with student progress or student development reached, ask the students. Ask the students how much progress they made. That accomplishes the first part of the problem to evaluate instruction. Evaluate instruction by asking the instructors what they are trying to do and asking the students whether or not they did it.

In the items 59-66 from Table Three you will note the item pool consists of one statement for each dimension. In this instance because group judgment is treated that is considered satisfactory. This results in a group average. When ends of that type are specified, one statement is probably satisfactory. If you want to go about student appraisal in this way several more items which ask these questions in various ways are needed. We are not trying to do that here. We are trying to find out about teaching effectiveness, which is a group judgment; therefore, one item is satisfactory.

Step Four. At this time trying out the items to see that they are satisfactory is required. Conduct some pre-tests, to find out if the students can do the items, or if they find items ambiguous. Test them in the sense that you compute some reliability co-efficients. That is the answer to the first question--evaluate instruction.

However, this problem was to evaluate instruction for purposes of improving it. When 'for purposes of improving it' was added, something besides items 59 to 66 was necessary. These items 1 to 58 are included in Table Four.

SEE TABLE FOUR: Student Reaction to Instruction and Courses: Items 1-58

Again go through the same process. Step One: What are the elements? In this case we are interested in teaching, in trying to improve instruction. We are not concerned with the elements of evaluation, but rather the elements of what goes on with the teacher. Teacher behavior and teacher attitudes are the major business or elements covered here.

Step Two: What are the dimensions of these characteristics? Again look up the literature to find out what people say, read treatises about what teaching really consists of, talk with colleagues, and review other instruments. The

TABLE FOUR

Student Reactions to Instruction and Courses
(Each item to be answered true or false)

1. The instructor seemed to have a well developed plan for each class session.
2. There were discussions between teachers and students. (As opposed to mere responses to questions.)
3. He explained course material clearly, and explanations were to the point.
4. The instructor seemed to lack energy.
5. The instructor answered students' questions as completely as reasonable.
6. He adjusted his pace to the needs of the class.
7. Class time was seldom or never wasted.
8. The instructor encouraged students to express themselves freely and openly.
9. He was often incoherent and/or vague in what he was saying.
10. The instructor seemed enthusiastic about the subject matter.
11. He generally spoke too rapidly.
12. He changed his approach to meet new situations.
13. On several occasions, he seemed unprepared for class.
14. Students made comments to the instructor without being asked.
15. He spoke with expressiveness and variety in tone of voice.
16. He demonstrated the importance and significance of his subject matter.
17. His presentations were dry and dull.
18. He requested and obtained student's questions and reactions.
19. He made it clear how each topic fit into the course.
20. He encouraged student comments even when they turned out to be incorrect or irrelevant.
21. He presented examples of what he wanted by way of homework, papers, etc.
22. He sometimes presented material in a humorous way.
23. He lectured in a low monotone.
24. He explained the reasons for his criticism of students' academic performance.
25. He failed to state clearly the course requirements and deadlines.
26. He attempted to induce silent students to participate.
27. He summarized material in a manner which aided in retention.
28. He stimulated students to intellectual effort beyond that required by most courses.
29. He lectured in a rambling fashion.
30. He understood student comments and questions even when these were not clearly expressed.
31. He stated clearly the objectives of the course.
32. He became angry or sarcastic when corrected or challenged by a student.
33. He failed to differentiate between significant and nonsignificant material.
34. He introduced stimulating ideas about the subject.
35. He repeated material to the point of monotony.
36. He displayed favoritism.
37. He related course material to real life situations.
38. He was available for individual help.
39. His speech was easy to understand.
40. He often dismissed class late.
41. He used leading questions to force students to answer their own questions.
42. He told the class when they had done a good job.
43. The examinations gave a balanced coverage to major topics.
44. The instructor gave ample notice for lengthy assignments.
45. The textbook (or substitute reading material) seemed out-of-date to me.
46. Too much of the course material repeated content covered by courses I had taken previously.
47. Examination questions were often unclear.
48. Out-of class assignments were reasonable in length.
49. The textbook (or substitute reading material) contained too little illustrative material.
50. Too much time was spent on too few topics-the course needed more breadth.
51. Examinations stressed memorization of information for which later recall seems unreasonable.
52. Assigned readings were pertinent to the topics presented in class.
53. Assigned readings (including text) were reasonably clear and understandable.
54. The instructor failed to make clear the relationship between this course and other courses.
55. Examination questions were frequently too detailed or picky.
56. I usually had no difficulty in obtaining outside reading materials.
57. Reading materials (including test) were organized in a logical, orderly fashion.
58. There were too many topics to understand any of them well.

product is a statement. We identified preparation and organization; student involvement, stimulation, speaking style, personalism, and consideration for students as the dimensions of these elements. Probably several more dimensions could be stated, but nothing is perfect. A start must be made somewhere. Consequently these items were written to tap those dimensions. (Item No. 1) The instructor seems to have a well-developed plan for each session, covers preparation and organization. There were discussions between teacher and students as opposed to mere responses to questions, gets at student involvement. So the item pool, the items 1 to 58 on Table Four were intended to cover these dimensions of the elements of teaching. To build this item pool we consulted students and teachers as to what they thought ought to be included.

The final step then is try out and test. In our test we administered this instrument in 700 plus classes where the instructor said gaining factual knowledge, objective No. 1 was an essential objective. Of the total group 200 or 300 classes identified this objective as important. Then it was a straightforward matter to check how did the students describe the teacher's attitudes and behaviors on items 1 to 58. In classes where the students report a lot of gain, did the students also report the instructor behaves differently or displays attitudes which are different from classes where very little gain was shown? Did the instructor seem to have a well-developed plan for each session? Did more students in classes where a lot of gain occured say True to that? If so, that would appear to be an item related to effectiveness in bringing about gain in a class. It is a relevant item. If there is no difference between degree of organization and planning that goes into a class, no difference in the amount of gain of factual knowledge--high or low--then that is not a relevant item for that particular objective.

Higher Education is coming to the point where this type of instrument will be used with increasing frequency. Students are almost forcing it, and by and large, it is a good thing. For too long incompetents have been rewarded and granted tenure. Students are aware of that, and they aren't going to put up with it much longer.

Question: Would you get into a problem now of trying to evaluate counseling in the very first part, trying to define what the effect of good counseling would be?

Dr. Hoyt: Counseling is a great illustration. Your purposes or your objectives change from client to client. To evaluate a counseling service, you have to bear in mind the fact that the objectives are somewhat different from one person to another. We have various types of information that must be distinquished. It is possible to gain information about occupations or occupational opportunities. This is a somewhat different task from gaining information about oneself. Even the area to gain information about one's talents or abilities, may be a little different from learning about one's interests or personal needs. These items get closer to areas which might be sensitive or more receptive to defensiveness. Therefore, parts on information are cognitive, are more or less cognitive, are factual and are parts of that which is emotional or more loaded emotionally. Distinctions of this sort must be made.

Question: Is this kind of statement all inclusive--the umbrella covering everything? Yet within that, can we look at some kinds of more specific things?

Dr. Hoyt: There is no one statement which has all these things in it. A whole series of statements is suggested so that information about occupations, information about training opportunities, information about abilities and talents,

information about interests and dispositions can be covered. These are separate sorts of things the client can respond to in terms of--Yes, I want a lot of that, it is of some importance, or it is not of importance at all. Reducing my personal discomfort, getting some sort of statement on that, feeling understood and appreciated, making a decision as to how to handle an interpersonal situation each item may be something to which the client should respond.

These would be dimensions of common elements that would constitute the purposes of your counseling interaction. These elements are partly cognitive, partly affective. There is the knowing, the feeling and the doing, or three elements which can be involved as far as counseling objectives go. For example, gaining information on occupations is a cognitive type objective. That would be one dimension. Occupational information or training opportunities would be a second dimension of that. Developing your plan of action with respect to educational career would be a motor or a doing dimension. Deciding how to handle an interpersonal situation would be another doing element. Reducing personal discomfort would be an effective dimension of an affective element.

The idea is to develop a reasonably comprehensive list. Specific items like learning how to deal with Uncle Charlie's drunkeness, cannot be incorporated. Somewhat more general terms must be used. Consequently learning to deal with a difficult interpersonal situation may be how the item is stated. Every such situation cannot be included. It is quite possible to develop a relatively brief list of objectives you would have along these elements--cognitive, affective, motor or doing. If the client's goals are the concern, the client should be in a position to say, "Yes, that is what I want to get done." Or say "that is what I am concerned about, and I want to make a plan, or I want to resolve this situation or I want to feel less tense, or whatever it might be. This is what I am at counseling for."

However, insofar as when purposes of counseling are left in terms of client needs, an interesting situation develops. These purposes may shift somewhat at times. The client enters and maybe purpose one is to get information about occupations, and his hidden purpose is to find out if you are or are not a straight shooter. If hidden purpose two is accomplished, hidden purpose three comes out which, for example, could be to learn how to deal with the difficult family situation. It is conceivable, therefore, that as one develops in this process, a more continuous accounting of expectations or felt needs must be envisioned with respect to the happenings during counseling situation. And so far as a given need was set to be very important, and its significance reduced, maybe then counseling was effective. This then is as far as you need to go in order to evaluate.

Question: How do we justify the use of the counselor's perception of the needs and then measure whether or not those needs were met by the client's responses afterwards? Perhaps the counselor does not recognize or overshoots the need. So when the client responds, they do not match up, we say the counseling has not been effective, or the needs are not met, while in fact perhaps the client's needs as he saw them were met.

Dr. Hoyt: There has to be some decision in advance to evaluate from what frame of reference. Is it from the reference that the client's needs were met, or from the frame of reference that the counselor's perception, that the counselor's objectives were met. In the long run it is a philosophical question as to what is it you are trying to evaluate. You have to be perfectly clear on it. You set up your evaluation study as to what you mean when you say--it is successful or it failed. At the same time quite possible that it succeeds from one

frame of reference but is a total failure from another. All the evaluation does is say you tell me what the frame is, and I'll tell you if it's yes or no, or some degree in between.

Question: Could you develop two different instruments? One to ascertain whether or not student's needs and goals were met, and one to figure out if your overall program needs were met?

Dr. Hoyt: Yes, there are various elements in evaluating an overall program; for example, an orientation activity which is not counseling at all but is introducing students to the campus making them feel a little more acquainted or comfortable. This activity may be part of your counseling program. Gaining effective study skills may be the part that is independent of this type of problem-solving or insightful program. Developing effective referral sources or gaining more acceptance for your service among faculty may be yet another type of counseling goal. If you have different thrusts you have to have different devices and ways for going about measuring. The one caution I would like to draw to your attention is, if your goal is to engage in a given activity, like to meet with faculty, or your goal is to talk with students, these goals have to be put in terms of having some sort of effect on somebody. Meeting with the faculty does not do anything to them. You can be successful in your goal then, but your program goes down the drain. Be very careful stating goals in terms of an activity you do. Concentrate on an effect. This frees you and sometimes makes you more creative by saying when writing goals, how can I best accomplish something.

RESEARCH PRINCIPLES

The last section of this report treats some of the principles involved in research, and some of the hang-ups which can interfere with the conduct of effective research. There are many kinds of research and almost all of it is useful. There is descriptive research where you simply are trying to get a picture of a situation, like a university or a sub-culture, to describe. There is predictive research where you are trying to say given this, what is likely to happen. This research can be useful, particularly in counseling. There is experimental research where the purpose is to discover some sort of causal relationship in terms of what makes what work. Research to prove, for example, that teaching French by method A is more effective than teaching French by method B, or to prove counseling had effects that no counseling does not. This research I will treat in some detail.

For a variety of reasons in education or in psychology experimental research is not done with complete finality. Many things can be done with a close to finality in fields such as agriculture and physical sciences. However, when you deal with human beings and human institutions, you are dealing with a lot of frailities which soil does not have. Nonetheless, having said that, I do not think we ought to give up. With a little logic it is possible to reduce plausible explanations as to why you got what you got down to a minimum. Then eventually you say the most logical reason we got what we got is that counseling did that, or that this teaching method did that, or that the atmosphere we created did that. It is possible to reduce those rival hypotheses to a minimum and discuss how that might be done. Interestingly enough, there are only three principles which govern real experimentation, but even those are hard to meet sometimes.

One is randomization. You must allow chance somehow to decide who will get a certain type of treatment and who will get a different type of treatment. In soil experiments you must decide by chance where this fertilizer is to be applied

or that fertilizer. The randomization so that chance enters into the experiment in a known way is important in experimentation.

You must have replication i.e. you must repeat whatever you do at least once. Partly because of the randomization which introduces a chance element, you cannot do it once and let it go. You must figure more than once in order to estimate how much is any variation in a treatment due to chance, and how much might be due to other factors--like counseling or like the fertilizer itself.

The third principle is you must have local control. There must be some plot of land to which no treatment was applied or some group of students who were not counseled or who did not get treated by this method.

If we were working under true experimental conditions, we would have to meet those three principles: randomization, replication and control. It is because of those rigid requirements that true experimentation is seldom possible in education. Frequently we must go to something called quasi-experimental designs where you do not have full control on randomization, where you do not have full access to a true control group.

In the use of those designs we introduce a number of factors which may jeopardize the validity of the study. That is, we introduce factors whereby it may be possible to account for the results of our study by something other than the particular treatment method in which we were interested. We may introduce factors that may explain the difference between a counseled and a non-counseled group on some basis other than counseling. This happens because we did not have full access to randomization or full access to a control group. And the problem then in planning such experimentation is to be aware of the factors which may jeopardize this type of validity or this ability to determine what the factor was. We must recognize these factors and plan the experiment in such a way as to minimize the possibility that such factors actually would operate. In other words put controls on these plausible rival hypotheses.

A few factors do jeopardize this type of validity. Two-thirds of the problem is knowing such things are there, and the other third is using that knowledge to arrange an experimental plan which will minimize these threats to the validity of your study. There are about eight such factors.

1) History: There may be some specific event beyond the experimental variable: treatment, counseling, teaching method, what not, some specific event or events which occur between the critical measurements. You measure the student beforehand; you measure him after the treatment has been applied. There may be some historical event that occurs between those measurements which has a noticeable effect upon the measure. For example, let's suppose we are interested in the effect of a reading accelerator machine on reading rate. At the beginning we give a reading test to determine student's level. He runs through a number of weeks of reading treatments, and then we want to give a second form of the reading test to see how much gain has occurred. Then the night before your post test is given, a campus riot takes place. This is an historical effect which very likely may jeopardize the validity of your experiment. One can expect that students will enter in a very agitated state of mind so much so that the true effect of the reading accelerator may be very difficult to determine under this condition. For example, one may be interested in the effect of homogeneous grouping policy on the question of school loyalty. Are students really more satisfied with their institutions and their education if they are grouped homogeneously on the basis of say their ability,

personality, or anything which in some way or other allows homogeneous grouping to occur. How students feel about the school is the chief interest. The president announces there will be a sharp fee increase of $100 next year. That is the type of historical event which can jeopardize the validity of your study. One must be aware of such historical events if valid experiments are to be conducted.

2) Maturation: This refers to any processes which operate within students or simply the passage of time "per se." Generally this means processes which occur as a function of growing older, but it can also mean processes which occur with the functions like growing hungrier, more tired. There has to be a concern about any process which may operate just as a function of time passing. A few years ago as a criterion in psycho-therapeutic research, it was popular to study shifts in self-ideal attitudes. Two sorts were used on this. The client was given a set of 100 characteristics, I am kind, patient, loyal, trustworthy, intelligent, clean shaven, etc. Then the client would sort these into categories of more or less descriptive of himself. When he finished that he had so many in the bottom category and so many in the top category and more in the middle and so forth. Then he would sort them again in terms of how he really would like to be. Those two sets would be correlated and that would be called a self-ideal correlation. Then we would test before psychotherapeutic experience and again after psycho-therapy to see whether those correlations between the self and ideal would shift as a function of psycho-therapy intervention. In the early studies of this type it was felt the true effects of psycho-therapy were noted because we got an increase in the relationship between self and ideal. It was not until later before this threat to validity was taken into account that people pointed out it may be a maturational effect. Studies were made on what changes occur without psycho-therapy just as a function of time, and, of course, when it was discovered those changes tended to be of the same order and magnitude as those associated with psycho-therapy, it looked like what was being displayed was a maturation effect, and not one effect of the treatment per se.

3) Testing threat: This refers to the effects of taking a test once upon the scores of a second testing. It is common, for example, in a program of remedial reading to give the Nelson-Denny initially, and then to give it again at the end of the remedial reading treatment. Then to attribute any differences, particularly gains, to the effectiveness of the reading program the student has undergone. The difficulty or one major difficulty with an experiment of that sort is it fails to take into account the fact that by taking the Nelson-Denny on two occasions, during the second occasion, the individuals have had the advantage of taking it the first time. Consequently, there will be something called the re-testing effect. It is not controlled under these circumstances, and it may be that the gains which are noted are not due, in fact, to the experimental reading program, but only to the retesting effect or the fact he took the test twice. There was a gain associated with that, and it is necessary then, if a person is interested in studying the effects of the reading program upon reading skills, to build in a control for effects of testing and re-testing on the same instrument.

4. Instrumentation: In some studies, there may be an instrumentation effect or instrumentation threat to the validity of your study. This refers to some changes in the calibration of the measuring instrument, such as changes in the observers or the raters...raters who are giving ratings which may produce some changes in measurement. That seems a little fuzzy, but maybe an example will clarify that. Suppose, we are concerned with the broad variable called the intellectuality of a student, and we wonder whether an experimental teaching program which is innovative, different from the usual lecture-recitation type thing might have some impact on the intellectuality of the students. There is no good

measure for intellectuality. We decide this has to be picked up on a judgmental criterion, and further we decide we will invite in the students in groups of three or four to talk about World Events, with a set of intellectual experts. These experts will make judgments on the degree of intellectuality displayed by the students. This will be done before our experimental program, and again afterwards. Where you are dealing with ratings of this sort, the difficulty lies in that the ratings which are accomplished after the program is over come from raters who are considerably more experienced than they were in the first place. When you started out in the pre-test, raters had not gone through this before. Now after they have interviewed a good many students and go about doing it again, they have had enough experience with it so they feel more comfortable. Perhaps they have changed the calibration of the instrument. Perhaps what they first said was low level, and now because their experience has changed a good deal what they see is medium level. Further, of course, if they know the nature of the experiment when it comes to interviewing experimental students as opposed to control students, there may be unintentional shifts in the calibration. Biases can influence changes in instrumentation. This threat to validity must be guarded against in your study.

5) Statistical Regression: Because in education we often deal with extremes, which is where this sort of phenomenon operates, this factor is particularly annoying and relevant. It operates where students have been chosen for a study because of their extreme position on some sort of scale. For example, suppose we were interested in the effects of a Creative Writing course on creativity scores. We are going to try to increase students' creativity by giving them a creative writing course. We give a pre-measure on creativity, and we choose to put into the creative writing course those students who scored in the lowest 20 percent in the creativity test. We use those for the experimental group, and those who scored above the twenty percentile in the control group. We don't give them the experimental creative writing course. Then we repeat the creativity test at the end of the experiment, and we compare the amount of gain made in creativity scores for the experimental and control groups. Under a condition like this we find that the experimental group will show a higher gain than the control group, even if nothing happened as a function of the creative writing course. And this happens because of statistical regression. For example, we give a measure, regents exams, and we take the lowest twenty percent on the total score and look only at total score. Unless the test has a perfect reliability 1.0, unless that happens, and it never has happened yet, that score the student has is an approximate estimate of his true score. It can be thought of as being made up of this true score--or where he actually does stand on the concept of being appraised. It is made up of that, plus an error of some sort. Because there are some errors in measurement the test is not perfectly reliable. It might be he misread a question, might be the question happened to sample a bit of the domain which he had not mastered. Even though possibly he had mastered a lot of the domain, we estimate he has not mastered very much because we happened to get the wrong question for him. It might be he put his response in the wrong place on the answer sheet. It might be he misread a word, there are many reasons why there is error in there, but there is error in all the scores. Some of the errors work to the advantage of the student, some of the errors work to his disadvantage. In the long run the positive and the negative errors balance out, but when we look at the total distribution of scores, we say that the mean is our very best estimate of our true mean.

If we are dealing with that lowest twenty percent, do the positive and negative errors balance out? No. The probability is if you are dealing with the lowest twenty percent, there are many more negative errors accumulated in that group. If you are dealing with the upper 20 percent, there are many more positive

errors in that group. What does that mean? If the student tries the test again, there is still chance as to whether the errors are going to help or hurt him. The first time the errors hurt him; the chances are he is not going to get hurt as badly the second time. The end result, his score is higher the second time, if to begin with, he was in the lowest twenty percent. What about the kids who began in the upper twenty percent? These students received help the first time with the errors working in their favor. The second time starting over, it is a new game. The phenomenon of extreme scores regressing toward the mean, statistical regression, will always occur when the reliability is less than perfect. The lower the reliability the more it will occur. That's the reason why this group who were low on creativity to begin with will have a higher score now due to statistical regression. It is a phenomenon of which we must be careful. Particularly this is true in our area, because we frequently choose as our subjects, students who are unusual or extreme in one way or another.

6) Selection: This threat refers to biases which may be introduced as a function of the selection of the comparison group. You have an experimental or treatment group. You want a comparison group with which to compare the results. The way the comparison group is selected may introduce biases which jeopardize the validity of your study. Suppose, for example, you are interested in the effect of counseling on social adjustment. And you chose for your study all students who score below the twentieth percentile on the Minnesota Test of Social Adjustment. Now, you have 150 kids who are chosen now because they were low on social adjustment as measured by this device. You invite them all by personal letter to come in and take advantage of your counseling services to deal with social adjustment problems. Sixty of them come in; ninety of them do not. You say the sixty are your experimental group. Now you have your control group. To begin with they are all alike because they are all low in social adjustment. That is an example of the way in which a selection bias might be introduced. And the bias, of course, is what did you control for, perhaps one of the more important factors, in determining whether social adjustment might change motivation to do something about it. Sixty students say "Yes, I'd like to take advantage of that." They appear to be quite different on the basis of motivation from the ninety students who ignored your letter. This is introduction of a selection bias and something which troubles us frequently in trying to set up valid experiments in our area.

7) Experimental Mortality: This refers to the differential loss of subjects in comparison to groups. This will be a problem to those working in community colleges where experimental or other types of mortality are very frequent. Mortality referring basically to whether they are still in your program. For example, you may be interested in, say of effect of an orientation course for low ability students on their grade point average. If you give a good orientation course aimed at "who am I," "what are my capabilities," "how can I take most advantage of them," etc. and you design that for low ability students, and you control the other factors discussed you end up with a group of low ability students who were allowed to take the course and another group which did not take the course. Divide it randomly--half and half--then you study the grade point at the end of the first term. You find there is a real difference in the grades earned by those assigned to the course and those not assigned. Is it possible to infer with validity then that the program orientation course did have the sorts of effects that were hypothesized for it? The course was an effective agent in increasing grade point average. In this instance you have to assume you controlled selection by suggesting to the students only so much room. Can you infer that? The experiment is threatened by the potentiality of experimental mortality. Let's say, you get a 1.9 average for your experimental group and a 1.5 for control. However 50 percent of the

experimental group are no longer there, whereas 90 percent of the control group are still enrolled. You may reflect this as due then to experimental mortality rather than to any effect of your program. One should be aware of this type of threat to the validity of a study and try to design studies so the ability to infer what it was one wanted it to infer is not jeopardized.

8) Selection Interaction Effects: This refers to interaction effects between selection and almost any of the other factors and may affect the validity of your study. An example of the way that would work: Suppose the question were to be the effects of experimental discussion program on the realism of vocational planning. This is what you wanted to discover, whether an experimental discussion program, you devised, would have any impact on the realism of vocational planning. You apply the treatment, the experimental discussion program, to sophomores enrolled at your institution. Compare then the realism of their vocational choices as judged by experts like yourself, with the realism of the vocational choices of freshmen. Under that condition you may have what is known as a maturation-selection interaction. You have selected, on the one hand sophomores, for the experimental group and freshmen for the control group. However, if the students mature in their development of realism of vocational planning at different rates, or there is a different amount of involvement already incurred in sophomores, then by this selection you have compounded the maturation explanation of your findings. A maturation and selection interaction has jeopardized the validity of that study.

A somewhat similar example might be the effect of group advising on attitudes toward the administration of your college. How do students feel about the administration--Dean, President, etc. You are going to see whether attitudes make them more favorable. If you apply this treatment, this group-advising program, to student government members as your experimental group, and your comparison-control group consists of leaders of religious groups of campus, it seems probably the experimental group will have held initially stronger attitudes toward the administration. This is true because student government leaders are naturally more directly involved with administration than leaders of student religious groups. If true that there were stronger attitudes, one way this would show up would be the reliability of the measure for student religious leaders would be lower. They do not care that much about administration; consequently they give you a different answer each time. Where the reliability for student government people would be considerably higher, this would be an example of a selection-testing interraction. These more or less complicated interactions of selection with other variables constitute the final threat to the validity of your study.

These are threats to studies, but they are not impossible problems. Your awareness of them makes it possible to do something about them, and to control them in some way to develop a very sound study. Knowledge of these eight points is as fundamental a knowledge to the conduct of effective research as anything that could be done. Certainly it is much more relevant than a knowledge of pre-tests, or analysis of variance or factor-analysis.

The "Hawthorne" effect, does not enter because these are threats to the internal validity of a study. Did something help or didn't it? The Hawthorne effect is an additional threat to validity, but it is an external threat. It affects the degree to which you can generalize to other situations and other schools. My point is really that this is the heart of good research or good studies.

There are many kinds of research all of which are valuable and helpful. Simple descriptive research projects, such as a profile of the entering freshman class, or a profile of the graduating students, or a profile of the students in a curriculum are valuable. What we know about student abilities and interests, their ambitions, and their perceptions of the educational environment provide helpful information. This data can be a source of some hypotheses, some reasoning, and common sense which is useful. However, where we are trying to improve the effectiveness of what we are doing this is not experimental research. Similarly, predictive studies, knowing about what will likely happen to students, are also very useful to counselors. Oftentimes you are faced with students' questioning their prospects. Counselors need to increase their knowledge about that, and be encouraged to conduct such applied research on a fairly routine basis.

BIBLIOGRAPHY

Adorno, T.W., E. Frenkel-Brunswik, D.J. inson, and R.N. Sanford. The Authoritarian Personality. rk: Harper and Brothers, 1950.

Astin, A.W. "A Re-examination of Col.. Productivity," Journal of Educational Psychology, 52: 173-178; 1961.

---------- "Some Characteristics of Student Bodies Entering Higher Educational Institutions," Journal of Educational Psychology, 55: 267-275; 1964.

---------- "Distribution of Students Among Higher Educational Institutions," Journal of Educational Psychology, 55: 276-287; 1964.

Cross, K. Patricia. The Junior College Student: A Research Description. Princeton: Educational Testing Service, 1968.

Davis, J.A. "Faculty Perceptions of Students, IV: Desirabilities and Perceptions of Academic Performance," Research Bulletin RB-64-13. Princeton: Educational Testing Service, 1964.

Ebel, Robert L. Measuring Educational Achievement. Englewood Cliffs, N.J.: Prentice-Hall, 1965.

Holland, J.L. and J.M. Richards, Jr. Academic and Nonacademic Accomplishment: Correlated or Uncorrelated? Research Report #2. Iowa City, Iowa: American College Testing Program, 1965.

---------- Academic and Nonacademic Accomplishment in a Representative Sample Taken from a Population of 612,000. Research Report #12. Iowa City, Iowa: American College Testing Program, 1966.

Hoyt, Donald P. The Relationship Between College Grades and Adult Achievement: A Review of the Literature. Research Report #7. Iowa City, Iowa: American College Testing Program, 1965.

Hoyt, Donald P. and L. Munday. Academic Description and Prediction in Junior Colleges. Research Report #10. Iowa City, Iowa: American College Testing Program, 1966.

Knapp, R.H. and J.J. Greenbaum. The Younger Adult Schoolar. His Collegiate Origins. Chicago: University of Chicago Press, 1953.

Trent, J.W. and L.L. Medsker. Beyond High School: A Study of 10,000 High School Graduates. Berkeley: Center for Research and Development in Higher Education, University of California, 1967.

Knoell, D.M. and L.L. Medsker. Factors Affecting Performance of Transfer Students from Two- to Four-Year Colleges: With Implications for Coordination and Articulation. USOE Cooperative Research Project No. 1133. Berkeley: Center for Research and Development in Higher Education, University of California, 1964.

Lehmann, Irwin J. Changes in Critical Thinking Ability, Attitudes, and Values Associated with College Attendance. USOE Cooperative Research Project #1646. 1962.

Lenning, Oscar, and Others. Student Educational Growth During the First Two Years of College. Washington: American Educational Research Association. 1968.

Lewis, C.S. Out of the Silent Planet. New York: Collier MacMillan, 1965.

Medsker, L.L. "The Junior College Student" in Junior College Student Personnel Programs: Appraisal and Development. A Report to the Carnegie Corporation. November, 1965.

Peterson, O.L., L.P. Andrews, R.S. Spain, and B.G. Greenberg. "An Analytical Study of North Carolina General Practice," Journal of Medical Education, 3., No. 12, 1956.

Telford, Charles W. and Walter T. Plant. Psychological Impact of the Public Two-Year College on Certain Non-Intellectual Functions. San Jose, California: San Jose State College, 1963.

---------- Personality Changes Associated with a College Education. USOE Cooperative Research Project No. 348. 1958-1962.

---------- Differential Personality Development in Young Adults as Related to Ability, USOE Cooperative Research Project No. S-042. San Jose State College. 1963.